



2810413668

Dispatched to Examiner

09/06/08 E



**Individuals, Society & the World:  
A Defence of Collective Environmental Duties**

**Elizabeth Blanche Cripps**

**University College London**

UMI Number: U593618

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U593618

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

### **Declaration**

I, Elizabeth Blanche Cripps, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

## **Abstract**

This thesis defends a collective duty to establish a global-level institution to tackle climate change. This is motivated through collective responsibility for environmental harm, and collective self-interest.

Building on Larry May, it is contended that a number of individuals can be collectively responsible, in a weak but morally significant sense, for the (even unintended) predictable result of the combination of their individual acts. It is argued that this gives rise to a collective duty to remedy or end the harm, and correlative individual duties.

The dominant intentionalist model of collectivities is rejected. Arguing against Margaret Gilbert, it is claimed that a collectivity is constituted by a set of individuals mutually dependent through some common goal, purpose or all-things-considered interest, whether or not they acknowledge it themselves.

A capabilities model of human flourishing is defended, according to which it is not in someone's all-things-considered interest to be deprived of a central functional capability, on something like Martha Nussbaum's list. To undermine a person's capability to enjoy a central functioning is to do morally significant harm.

It is argued that, especially for larger collectivities, it is often "better", in terms of achieving the common goals, purposes or interests, that certain decisions be made collectively, rather than left to the aggregation of individual acts. This appeals to: inefficiency, ignorance, the individual-collective rationality distinction, partial conflict, and rational altruist arguments.

Collective (and correlative individual) duties to establish global environmental decision-making institutions are defended: prudentially, because most humans constitute a collectivity by virtue of the threat of climate change to fundamental interests, and morally, because most are collectively responsible for harm.

Finally, institutional change is called for, so that certain (primarily environmental) decisions are made by a global decision-making body and handed down as restrictions on states and individuals. A number of objections are addressed.

## Contents

|   |            |
|---|------------|
| <b>Acknowledgements</b>   | <b>6</b>   |
| <b>Introduction</b>   | <b>7</b>   |
| <b>Chapter One: The Harm We Do Together</b>   | <b>12</b>  |
| I. A “gap” in conventional morality   | 14         |
| II. May’s collective responsibility   | 16         |
| III. Weak collective responsibility   | 20         |
| IV. Intentionality and overdetermination  | 26         |
| V. Collective and individual duties   | 28         |
| <b>Chapter Two:<br/>Against the Intentionalist: Collectivities v. Social Groups</b> | <b>36</b>  |
| I. The non-intentionalist model   | 38         |
| II. Gilbert’s intentionalist model  | 42         |
| III. Three objections to Gilbert  | 45         |
| IV. Accommodating tough cases   | 51         |
| V. Defending the distinction 1: social properties, groups and aggregates            | 53         |
| VI. Defending the distinction 2: collective v. social properties                    | 55         |
| VII. Avoiding circularity   | 57         |
| VIII. Controversial collectivities  | 58         |
| IX. Meeting Graham’s conditions   | 60         |
| X. Two objections to the non-intentionalist model                                   | 68         |
| XI. Morally salient potential collectivities  | 72         |
| <b>Chapter Three: In the Human Interest: Defending Capabilities</b>                 | <b>75</b>  |
| I. Human flourishing: the capabilities approach                                     | 77         |
| II. The Dworkinian complaint 1: resources & preference satisfaction                 | 83         |
| III. The Dworkinian complaint 2: welfare & paternalism                              | 88         |
| IV. More than preference satisfaction   | 93         |
| V. Paternalism: acceptable and unacceptable   | 98         |
| VI. A final objection   | 105        |
| <b>Chapter Four: Collectivities as Units of Agency</b>                              | <b>109</b> |
| I. Better for whom?   | 110        |
| II. Inefficiency arguments and the Discursive Dilemma                               | 113        |
| III. Ignorance arguments  | 120        |
| IV. Game theory arguments   | 122        |
| V. Partial conflict arguments   | 128        |
| VI. Rational Altruist Argument A: Parfit’s classic case                             | 130        |
| VII. Rational Altruist Argument A: imperceptible benefits                           | 133        |
| VIII. Rational Altruist Argument B: only a chance of benefiting                     | 140        |
| IX. Rational Altruist Arguments A&B: changing the unit of agency                    | 142        |
| X. Clarification and reassurance  | 145        |

|  |            |
|--|------------|
| <b>Chapter Five: Environmental Harm and Global Collectivities</b>        | <b>151</b> |
| I. Self interest and the environment                                     | 152        |
| II. A prudential “ought”: collective rationality                         | 158        |
| III. A moral ought: the Principle of Fairness                            | 160        |
| IV. Motivating George: duties of humanity                                | 163        |
| V. Environmental harm and moral excuses                                  | 169        |
| VI. Rights and duties  | 178        |
| VII. Towards legitimate coercion 1: collectivising the Harm Principle    | 179        |
| VIII. Towards legitimate coercion 2: the Principle of Fairness revisited | 184        |
| <b>Chapter Six: Implications and Objections</b>                          | <b>192</b> |
| I. The Decision-Making Claim   | 194        |
| II. Weakening state supremacy  | 198        |
| III. Objections to the Decision-Making Claim                             | 200        |
| IV. Same benefits argument: is state co-operation enough?                | 204        |
| V. Moral burdens argument  | 211        |
| VI. Past harmers objection   | 217        |
| VII. Too inclusive objection   | 218        |
| <b>Conclusion</b>  | <b>224</b> |
| <b>Bibliography</b>  | <b>227</b> |



## Acknowledgements

I would like to thank the Arts and Humanities Research Council and the Royal Institute of Philosophy for the funding which made this project possible.

Much of the work in this thesis has been presented at academic conferences and I have benefited greatly from the feedback received. Earlier versions of Chapter One and Sections V, VI and VII of Chapter Five were presented as “Collective Harm and Environmental Duties” at: Brave New World, Manchester University, 27-8 June 2007, the Manchester Political Theory Workshops, 4-6 September 2007, and the V Pavia Graduate Conference in Political Philosophy, 6-7 September 2007. I am also grateful for insightful comments on the paper from many individuals, including Harry Cripps, Jennifer Clare Heyward and Tom Porter.

Chapter Two draws extensively on my 2005 MPhil thesis, *Individuals and Society: A Non-Intentionalist Account of Collectivities and Social Groups*, for which I received careful and helpful supervision from David-Hillel Ruben. I am also grateful for feedback from my MPhil examiners, Patrick Riordan and Fabienne Peter.

Chapter Three has gained from responses to my paper, “Defending Capabilities: Objectivity without Unacceptable Paternalism”, at the PhD workshop at the Department of Political Science, UCL, on 14 March 2007, and the Human Development and Capabilities Association Conference at the New School, New York, 15-20 September 2007. I am also grateful for comments on this paper from Gry Wester.

Chapter Four, Section II draws on my paper, “The Discursive Dilemma and Collective Decision Making”, given at the Institute of Philosophy London Graduate Conference on 10 February 2006. A version of the overall chapter was presented to UCL Philosophy graduates and European Political and Social Studies undergraduates at a workshop in UCL on 26 February 2008.

Finally, I am grateful to Laura Valentini for comments on Chapter Six, to Michael Otsuka for detailed and helpful discussion on several chapters, to Vivien Cripps for proof reading, and especially to Jonathan Wolff for advice, encouragement and constructive criticism throughout the entire writing process.

## Introduction

*It's exciting to have a real crisis on your hands, when you have spent half your political life dealing with humdrum issues like the environment.*

Margaret Thatcher, 1982<sup>1</sup>

*[T]he scientific evidence of global warming caused by greenhouse gas emissions is now overwhelming... [T]here is nothing more serious, more urgent or more demanding of leadership, here of course but most importantly, in the global community.*

Tony Blair, 2006<sup>2</sup>

As a species, we face urgent global problems.<sup>3</sup> Many of these result from human behaviour. The detrimental impact of combined human activity and human numbers on the natural world is beyond dispute, although the precise levels, the feasibility of prevention and the danger of passing a “threshold” beyond which damage becomes exponentially worse and irreversible, remain open to scientific debate.<sup>4</sup> These are problems which will affect us all but which have, and will continue to have, a particularly detrimental effect on some of us, especially those who are already disadvantaged in various ways.<sup>5</sup>

Through this impact on the environment, although not only in this way, we can seriously influence the lives of ourselves and others on a worldwide scale. Thus, like it or not, we find ourselves in an inescapable web of relations extending beyond the acknowledged “groups” within which each of us sees herself as living her life: beyond such familiar categories as families, friendship groups, local communities, nation states and even federal groupings of states. Our impact on the lives of others extends, moreover, beyond the current generation.

---

<sup>1</sup> Margaret Thatcher, then British Prime Minister, on the Falklands War, to the Scottish Conservative Party Conference, 14 May 1982. (*Oxford Dictionary of Quotations*, Revised Fourth Edition, 1996. Exact wording sourced by the *ODQ* from Hugo Young: *One of Us* (Macmillan, 1989)).

<sup>2</sup> Tony Blair, then British Prime Minister, at the launch of the Stern Review into climate change, 30 October 2006. (Full speech available at [www.number-10.gov.uk](http://www.number-10.gov.uk))

<sup>3</sup> Indeed, “we” do so more generally – as all the species sharing this finite planet.

<sup>4</sup> See, for example, IPCC 2007 and Orestes 2004. For a survey of scientific and philosophical arguments, see Gardiner 2004, 555-600. For his discussion of threshold effects, see 2004, 562-3. See IPCC 2007, 53-4 for analysis of the risk of abrupt or irreversible change.

<sup>5</sup> See IPCC 2007, 48-53 and Shrader-Frechette 2002, 8-13

These are practical points, but they raise pressing questions for moral and political philosophers – those thinkers who concern themselves with how human beings ought to live, and how they ought to live *together*. The philosophical challenge is all the greater given that the nation state has so long been taken as the locus of political responsibility, and moral philosophy has so long focused on the foreseeable impact of our *individual* actions (or the action of some clearly defined corporate body).

The kind of questions facing us are: What should we (prudentially and morally) do about climate change? What does fulfilment of our duties require in terms of a global institutional restructure? What does this mean for each of us, as an individual? Under what circumstances could environmental duties legitimately be coercively imposed? To answer them, we need to go back to the moral philosophy and philosophy of social science on which political philosophy relies, to ask some still more basic questions. We need to reconsider the circumstances under which we constitute the kind of group or collectivity with which political philosophy should be concerned, and those under which we can be criticised as doing morally significant harm to others.

This thesis will not answer all the questions posed above, but it will make a start. It will defend a collective duty to establish a global environmental decision-making body, and will do so in two ways: prudentially, by appeal to collective self-interest, and morally, through collective responsibility for environmental harm. It will also discuss correlative individual duties: to promote such an institution (assuming there is a chance of success) and (subject to various conditions) to comply with the restrictions it imposes. These are motivated both prudentially and morally. In the moral case, supplementary individual duties are also suggested.

I will argue that such a global institution, to be effective, would need powers of enforcement. However, that is not enough in itself to justify such coercive authority. A first step will be taken towards such a justification, initially by appeal to the prudential incentive to consent to mutual coercion, but supplementing this with two central liberal ideas: the Harm Principle and the Principle of Fairness, the latter itself supplemented by appeal to some positive moral claims, or duties of humanity.

The thesis will be structured as follows. Chapter One will build on Larry May's work to defend a weak but important notion of collective harm. The claim is that a number of individuals can be collectively, morally responsible for the predictable result of the combination of their individual acts, *even if* they do not individually intend to contribute to harm, or even to some collective act. I will argue that this weak collective responsibility gives rise to a collective duty to remedy or end the harm (if necessary by establishing collective decision-making structures), and also suggest some correlative individual duties.

Chapter Two will reject the dominant intentionalist model of collectivities, or social groups, engaging in particular with Margaret Gilbert's plural subject account. (The "intentionalist" claim, broadly put, is that group-wide acknowledged awareness of membership is necessary for collectivityhood.) Instead, I will argue that a collectivity is constituted by a set of individuals mutually dependent through some common or shared interest, purpose or goal, and that *social groups*, on which political philosophers and philosophers of social science have standardly focused, are only a subset of the wider (and politically and explanatorily relevant) set of collectivities.

The mutual dependence clause is understood broadly: as meaning that the individuals have at least one goal, purpose or interest that it only makes sense to consider as being pursued together, *whether or not they acknowledge it themselves*. I will also introduce the idea of a morally salient potential collectivity, whose members would constitute a collectivity were they to espouse some particular goal, and who have a collective moral duty (because of collective responsibility for unacceptable harm) to espouse it.

Assumptions made in both these chapters will be filled out by Chapter Three. I will defend a capabilities model of human flourishing, according to which it is against someone's interest (whether or not they acknowledge this) to be permanently deprived of a central *capability to function*, on something like Martha Nussbaum's list. By the same token, to undermine a person's capability to enjoy one or more central functionings is to harm them in a morally significant way.

Chapter Four will explain how a collective moral duty (or collective rational incentive) can translate to a duty to set up a collective decision-making structure. It will do so by defending the claim that, especially for larger collectivities, it can

be better for certain decisions to be made at the collective level, rather than left to the aggregation of individual acts. “Better” is understood strictly in terms of achievement of the goals, purposes or interests around which a collectivity is constituted. I will appeal to: efficiency; individual refusal to acknowledge common interests; the individual-collective rationality distinction; partial conflict; and two rational altruist arguments.

Chapter Five will apply the theoretical framework established thus far to the environmental case. It will argue that most humans make up a collectivity by virtue of a fundamental interest in environmental preservation, and that a slightly different group, also constituted by much of humanity and including any individuals not in the first set, constitutes a morally salient potential collectivity through weak collective responsibility for environmental harm. By dismissing potential moral “excuses”, I will argue that this harm is morally impermissible.

This, I will contend, gives us a collective duty to establish a global environmental decision-making institution. Individual duties to promote and comply with such an institution will also be defended, and a preliminary argument will be offered as to how coercive powers of enforcement, by such an institution, could be legitimated.

Finally, I will outline some institutional implications. Building on Chapter Five, I will specify a Decision-Making Claim, according to which decisions should be made *qua* collectivity when they are directly related to achieving or satisfying the common purposes, interests or goals of the collectivity, *and* need to be made collectively to achieve better those goals, purposes or interests. This will be used to defend the making of environmental decisions by a global institution, and a corresponding weakening of state sovereignty. I will also engage with potential criticisms of both this proposed institutional framework (particularly statist objections to the threat to state sovereignty) and the philosophical model on which it rests.

I am aware that my account will prompt criticisms from all sides, breaking, as it does, with many conventions in this overlap of three philosophical fields: moral philosophy, political philosophy and the philosophy of social science. I am unapologetic. If we are required to stick to a framework which provides no resources for the political philosopher to address the results of our *practical*

connectedness and the combined impact of our behaviour (which does not always follow institutional lines but, rather, may motivate institutional change) then we are doomed, as far as policy making is concerned, to practical and moral disaster.

## Chapter One

### The Harm We Do Together

As a result of (largely anthropogenic) climate change, future generations will inherit a world very different from that into which we were born. Temperatures will be higher, weather patterns will be more extreme, an increasing proportion of the currently habitable world will be under water, and the rest will be at greater risk of floods, tropical cyclones, droughts or other natural disasters.<sup>1</sup> The statistics are bleak even for the lifetimes of those currently alive.<sup>2</sup> Those in low-lying or otherwise vulnerable parts of the world are already affected by changing weather patterns. Pollution and other environmental waste is having a dramatic negative impact on health in poor communities even in otherwise affluent states, as well as in developing countries.<sup>3</sup>

What should this mean to me, as an individual enjoying a lifestyle which may not be carbon-intensive compared to that of many of my fellow westerners, but is certainly so compared with much of the world's population? Should I refrain from flying to the Costa Blanca for a weekend's rock climbing? Should I refrain even from driving to Snowdonia? Should I use less paper, glass and plastic? Should I recycle what I do use? Should I petition the UK government (or the EU parliament) daily to change its climate change policy? Should I spend every conversation with friends or colleagues nagging them to use less energy? Should I, as well as or instead of all these, give money to environmental charities, or to

---

<sup>1</sup> Average temperatures are expected to increase by 0.2°C per decade for the next two decades. After that, estimates suggest a 0.1°C increase per decade even if concentrations of greenhouse gases were maintained at 2000 levels. Average temperatures in 2090-9 could be up to 6.4°C higher than in 1980-99, depending how pessimistic a forecast one takes on emission growth. Sea levels are predicted to rise by up to 0.59m. (IPCC 2007, 45-54) Beyond the 21<sup>st</sup> century, the IPCC warns: 'Anthropogenic warming and sea level rise would continue for centuries due to the time scales associated with climate processes and feedbacks, even if GHG concentrations were to be stabilised.' (46) Gardiner, drawing on US National Research Council statistics and Robert Gagosian: "Abrupt Climate Change: Should We Be Worried?" (Woods Hole Oceanographic Institution, 2003) warns that IPCC predictions may not be stark enough, as they could underestimate 'the potential for nonlinear threshold effects'. (2004, 562)

<sup>2</sup> IPCC 2007, 50-3. In Africa, the report warns: 'By 2020, between 75 and 250 million of people are projected to be exposed to increased water stress due to climate change. By 2020, in some countries, yields from rain-fed agriculture could be reduced by up to 50%. Agricultural production, including access to food, in many African countries is projected to be severely compromised. This would further adversely affect food security and exacerbate malnutrition.' (50) The IPCC also asserts with '*medium confidence*' that the impact of temperature changes has *already* been felt in 'some aspects of human health'. (2007, 33)

<sup>3</sup> Shrader-Frechette 2002, 8-13

charities helping the victims of environmental disasters? (If so, how much? Any spare cash after fulfilling my wants, or all bar what is needed to keep me above some sufficiency threshold?<sup>4</sup>)

I might query whether I should do any of these things, given that my own actions are not directly responsible for any particular harm. My turning up the heating will not make any perceptible difference to the temperature of the planet, pollution levels in Mexico City, or the size of the rainforest. Nor will my turning it *down*, in itself, make any positive difference. So, if I think only in terms of the consequences of my individual actions, I will not change my way of life.

Nonetheless, it is as the result of the combined behaviour of millions of people like me that such damage is being done to the planet, and so to human beings.<sup>5</sup> Perhaps, simply on this basis, I should refrain from flying (or driving) at all, or from heating my house above the minimum? (I would then be appealing to the Kantian idea that I should not do anything that I cannot consistently will that everyone should do: if everyone lived the high energy lifestyle of the average Londoner, the result would be catastrophic.<sup>6</sup>)

However, my sacrifice would be pointless unless others also changed their behaviour. If, by refraining from *all* CFC-generating activities, I did make some significant difference, I would (surely unfairly) be bearing the costs of others' profligacy.

This chapter will introduce a framework within which the questions posed above can appropriately and helpfully be considered. It will do this by filtering our individual duties in such cases through an expanded notion of moral responsibility for harm, according to which the primary responsibility (and so the primary correlative duty) is *collective*. For an individual, the relevant question is not simply, "What should *I* do?", independently of how others could change their

---

<sup>4</sup> Singer, for example, would claim that each of us has this strong duty, simply as a duty of humanity. (1972, 229-43)

<sup>5</sup> Note that I base the arguments of this thesis on the harm done to human beings – present and future. However, this should not be taken as a claim that the flourishing of humans is the *only* thing that matters. My post-doctoral research proposal includes the expansion of my model to include the unacceptability of environmental harm to non-human animals, and I would not rule out a moral duty based on the intrinsic value of the natural world.

<sup>6</sup> Singer 2002, 35. Kant's categorical imperative, in its first formulation, requires each individual to 'act only in accordance with that *maxim* through which you can at the same time will that it become a universal law'. (Kant 1785, 31, emphasis in original) Thanks to Andrew Williams for bringing this point to my attention.



behaviour, but rather: “What should *we* do, and what duties do I acquire, through that collective duty?”

I will defend a notion of *weak collective responsibility*, according to which a set of individuals can be collectively responsible for harm caused by the predictable aggregation of individual acts, and explore its implications for collective and individual duties. Chapters Five and Six will apply this to the current environmental scenario.

This chapter will be structured as follows. Section I will motivate my investigation of collective responsibility by appeal to Samuel Scheffler, who highlights a “gap” in our standard moral thinking. I will then engage with May to defend an expanded understanding of collective responsibility (in a weaker sense than corporate responsibility). After outlining May’s arguments, laying out my model of weak collective responsibility and distinguishing it from his account of shared responsibility, I will defend the notion of weak collective responsibility by rejecting two potential objections. (Sections II, III and IV) Section V will consider the implications of the model in terms of collective (and correlative individual duties) and complete the defence of the account by rebutting a final objection.

## **I. A “gap” in conventional morality**

Scheffler, appealing to the often unintended and unacknowledged global “processes” to which our individual actions contribute, draws attention to a possibility overlooked by standard moral thought: moral duties that bind a number of individuals *collectively*, even though it would be inappropriate to *blame* them individually and they are not acting in combination through some social or institutional structure.<sup>7</sup> It is evident, in practice, that such groups (in this very weak sense) can and do cause great, and morally regrettable, harms.

[D]evelopments in the modern world... [including] ... remarkable advances in science and technology... continuing revolutions in travel, communications, and information processing... increased economic and political interdependence among the countries of

---

<sup>7</sup> Scheffler 2001, 32-47

the world: and the enormous growth in world population... have made it more difficult than ever to sustain the conception of human social relations as consisting primarily in small-scale interactions among single individuals... [Q]uality of life for people in any one part of the world is, to a very large extent, a function of a network of institutional arrangements that supports a very different quality of life for people in other parts of the world... [P]olitical and economic developments in one area of the globe often have rapid and dramatic effects on people in other areas... Moreover... human behaviour now has effects on the natural environment that are unprecedented in scale. These environmental effects distribute themselves in complicated ways within and across national boundaries...<sup>8</sup>

Scheffler suggests that we find such scenarios difficult to accommodate because we tend to think exclusively in terms of *individual* moral responsibility, within which realm the so-called common sense view has a strong intuitive hold on us. This view centres on two conceptions of normative responsibility: ‘the doctrine that individuals have a special responsibility for what they themselves do, as opposed to what they fail to prevent’, and the idea that ‘one has distinctive responsibilities – or “special obligations” – towards members of one’s own family and others to whom one stands in certain significant sorts of relationships’.<sup>9</sup>

Thus, he argues, it would be problematic on two grounds to try to accommodate the global situation whilst retaining an exclusive focus on *individual* moral responsibility. Firstly, practically, to address these global problems with only the vocabulary of individual responsibility would mean rejecting the common sense view – one which has a deep hold on us. Secondly, more fundamentally, while these political and economic developments raise cases on which the narrow, common sense approach is ill-equipped to comment, they do not automatically provide an alternative within the framework of individual moral responsibility.<sup>10</sup> ‘[I]t is not uncommon for an important outcome to be the product of a large number of acts performed by many different

---

<sup>8</sup> Ibid., 40

<sup>9</sup> Ibid., 36

<sup>10</sup> Using a parallel argument to those against consequentialism, to attempt to account for such wide ranging effects within the framework would mean putting unreasonable demands on individuals. ‘After all, the individual agent *qua* individual agent will typically have only the most limited opportunities to influence these global dynamics, and, indeed, cannot in general be assumed to have any but the sketchiest and most speculative notions about the specific global implications or his of her behaviour.’ (Scheffler 2001, 43)

people, few if any of whom actually intend to produce the outcome in question.’<sup>11</sup>

Thus, according to Scheffler, acknowledging the moral salience of a global perspective does not necessitate a re-think of the conception of individual responsibility but rather ‘generate[s] doubts about the practice of treating the individual agent as the primary locus of [normative] responsibility’.<sup>12</sup> This chapter will make the obvious next move, exploring the notion of collective responsibility as the best way to fill this hole in our moral framework.

That is *not* to say that it will ignore the question of individual duties (of what *I* should do) altogether. The claim is, rather, that individual duties in such cases are most appropriately filtered through the question of what harm *we*, collectively, are responsible for, and our corresponding *collective* duty to do something about it. Individuals, on my model, acquire duties, or responsibilities, to behave in certain ways. However, they acquire them as part of a group (in some weak sense) which is collectively responsible for some harm. Most crucially, the individual is not required to think only in the restrictive terms of whether she should perform some particular contributory act, but in terms of how (if at all) she can change what is happening at the collective level.

## II. May’s collective responsibility

This section will provide a brief outline of May’s model of collective responsibility. It will also summarise his account of *shared* responsibility, which he uses to accommodate some of the cases where, I will argue, my weaker collective responsibility also applies. The following sections will build on his arguments to present and defend my own account.

According to May, if there is *collective* responsibility for a harm, some group as a whole is responsible: it is not necessary that all or any members be individually so. Collective responsibility, he adds, concerns ‘the nondistributed

---

<sup>11</sup> Scheffler 2001, 43

<sup>12</sup> Ibid., 44

responsibility of a group of people structured in such a way that action can occur that could not occur if the members were acting outside the group'.<sup>13</sup>

A relatively familiar alternative to individual responsibility is that of corporate responsibility, which seems to fit naturally into this understanding of "nondistributed" responsibility. Corporate responsibility assumes a formalised decision-making apparatus, because of which there is a sense in which the corporation can be said to have acted which is not reducible to the contributing acts of individuals. The corporation can be held responsible as a unit (thus, the *company* Network Rail was fined £14m by the Office of Rail Regulation;<sup>14</sup> the UK *government* is popularly blamed for taking part in George Bush's invasion of Iraq in the face of inadequate evidence of weapons of mass destruction) although that does not preclude assigning *personal* responsibility (for all or some part of the harm) to certain individuals within the corporation (for example, Tony Blair).

However, May extends the idea of collective responsibility to include more than the standard corporate case.<sup>15</sup> Firstly, he attributes collective responsibility for harm to informal groups such as mobs, where there can be, he maintains, group action.<sup>16</sup> He argues that, '[a]lthough mobs do not have explicitly decision-making structures', there is a sense in which all members contribute to some action, and there is 'a kind of collective intention'<sup>17</sup> by virtue of mob solidarity.

Secondly, May defends the idea of collective responsibility for *inaction* by a "putative group" – that is, where no group existed but one could have been formed in time to prevent a harm from taking place.<sup>18</sup> His example is of a

---

<sup>13</sup> May 1992, 106

<sup>14</sup> The fine was imposed after a four-day overrun on engineering works on the main London-Glasgow railway line. (See, for example, *The Guardian*, 29 February 2008)

<sup>15</sup> May 1987 and 1992. Others have also expanded the model of collective responsibility. David Miller, for example, defends an account on which either "like-mindedness" or participation in and benefit from a "cooperative practice" is sufficient for collective responsibility, in the sense of responsibility to remedy. (2004, 248-57) However, he does not rule out a wider model: 'I don't want to claim that it is a necessary condition for ascribing collective responsibility to a set of individuals that they should display like-minded group or cooperative practice features.' (2004, 249)

<sup>16</sup> 'I use the term "mob" to refer to social groups which have no decision-making or organisational structures, but which, unlike random crowds, have come together for economic or political reasons.' (May 1987, 73-4)

<sup>17</sup> May 1987, 74

<sup>18</sup> "Collective *inaction*", according to May, refers to 'the failure to act of a collection of people that did not choose *as a group* to remain inactive but that could have acted *as a group*', as opposed to "collective *omission*", which 'refers to the failure of a group that collectively chooses not to act'. (1992, 107)

number of holiday makers on a beach, who fail to organise themselves in time to save a child from drowning.<sup>19</sup> Collective responsibility in such cases hinges, he argues, on whether the individuals could have organised in time to prevent the harm.

A putative group of people engages in the kind of inaction that warrants collective responsibility if:

- (a) the members of the group fail to act to prevent a harm, the prevention of which would have required the coordinated actions of (some of) the members of the group;
- (b) it is plausible to think that the group could have developed a sufficient structure in time to allow the group to act collectively to prevent the harm; and
- (c) it is reasonable to think that the members of the group should have acted to prevent the harm rather than doing anything else, such as preventing other harms which they also could have prevented.<sup>20</sup>

May argues that while the individual members of the putative group should not generally be blamed or feel guilt, ‘feelings of shame or taint’<sup>21</sup> are appropriate.<sup>22</sup>

May also defends a model of *shared* responsibility. This, unlike his understanding of collective responsibility, ‘concerns the aggregated responsibilities of individuals, all of whom contribute to a result and for that reason are *personally* responsible, albeit often to different degrees, for a given harmful result’.<sup>23</sup> However, shared and collective responsibility are not mutually exclusive. Rather, May argues that a number of individuals can share personal responsibility and still be collectively responsible in the nondistributive sense for bringing about some harm.<sup>24</sup>

May develops his notion of shared responsibility by appeal to cases where the *joint action* of two or more people leads to harm. In joint undertakings, he argues, ‘the causal contribution of each person often cannot easily be ascertained,

---

<sup>19</sup> May 1992, 110

<sup>20</sup> Ibid., 116-7

<sup>21</sup> Ibid., 109

<sup>22</sup> Both the formalised corporate structure and the mob would count as collectivities on the model I draw up in Chapter Two. A putative group would count as a potential collectivity. However, I will not use this vocabulary until my model has been laid out.

<sup>23</sup> May 1992, 107 [emphasis added]

<sup>24</sup> Ibid., 107. He uses the example of the downing of an Iranian passenger plane by the US military: ‘Collective responsibility was attributed to the United States; but in addition, certain Americans, because of their participation, shared responsibility for what occurred.’ (107)

except to say that all parties played a necessary role in the harm, and that no one party played a sufficient role'.<sup>25</sup> He goes on to treat joint action as synonymous with *concerted* action, which he defines around intentionality.

Concerted action requires... some sort of tacit understanding or agreement among the parties. But such an understanding is not reached when there is "mere knowledge by each party of what the other was doing". Rather, in addition, there must be one clearly recognised goal (or set of goals) toward which each sees their actions as directed.<sup>26</sup>

His example is of a number of persons who, between them, stole some silver buttons. One battered down the door, another restrained the plaintiff, another took the buttons, and so on, all intentionally contributing to the overall result.<sup>27</sup>

Although he focuses on intentional cases, May indicates two further scenarios in which 'shared rather than individual full responsibility seems appropriate'. The first is that in which a person acted (voluntarily, even intentionally) 'with no knowledge or intent of contributing to a harm [but nonetheless] did perform a necessary part in a harm'. The second is when 'a person... did not directly cause a harm [but] nevertheless acted so as to increase the risk of the occurrence of harm'.<sup>28</sup>

The idea of shared personal (individual) responsibility naturally gives rise to the question of how much harm each individual is personally responsible *for*. May suggests three ways in which personal responsibility might be allocated in such cases. The first is division of responsibility: either according to different roles, which could be used to assign responsibility for proportionally extensive or limited "parts" of the harm, or by a straight division of the harm among the individuals. The second possibility is that no individual takes responsibility. Thirdly, each takes full responsibility.<sup>29</sup>

However, the latter two are often inappropriate. Given, May argues, that the participation of each individual is necessary for the harm, *some* responsibility is surely appropriate. But because no individual's actions, alone, would have been

---

<sup>25</sup> May 1992, 39

<sup>26</sup> Ibid., 41

<sup>27</sup> Ibid., 40. He borrows the example from W. Paige Keaton ed. *Prosser and Keeton on the Law of Torts* (West Publishing Company, 5<sup>th</sup> edition 1984), 322-3

<sup>28</sup> May 1992, 39

<sup>29</sup> Ibid., 42

sufficient to cause the harm, full responsibility seems inappropriate.<sup>30</sup> Thus he favours the first option and, within that, dividing responsibility according to the roles played by individuals in bringing about the harm.<sup>31</sup>

### III. Weak collective responsibility

The previous section outlined May's expansion of the idea of collective responsibility beyond the narrow corporate model. However, I contend that even his account overlooks a further plausible understanding of collective responsibility: a weak, but hugely significant, sense in which a number of individuals can be responsible for the combined result of their individual actions. The next sections will explore this possibility, distinguish weak collective responsibility from May's shared responsibility, defend it against potential objections and discuss its implications in terms of collective and correlative individual duties.

The claim to be defended is:

#### *Weak Collective Responsibility Claim*

A number of individuals who do not yet constitute a collectivity (either formally, with an acknowledged decision structure or informally, with some vaguely defined common interest or goal, as with the mob) can be held morally *collectively* responsible for a harm which has been caused by the predictable aggregation of individual actions.

"Predictable" here has a double meaning: individuals are acting predictably, in that they are acting in pursuit of their goals, purposes or interests, and it is predictable (reasonably foreseeable) that those actions, in combination, would result in the harm.

---

<sup>30</sup> Ibid., 39

<sup>31</sup> 'I prefer the uneven-division variation of this strategy.' (May 1992, 42) In collective *inaction* cases. May argues that each individual's share of responsibility will depend on the role she could have played in bringing about a collective prevention of the harm (for example, on her leadership skills if such skills would be crucial in enabling the putative group to act as a group). (1992, 117)

Consider the following situation, a variant of May's collective inaction example.<sup>32</sup> A number of teenagers, all independently, decide to swim in a small lake. They dive in, swim around very flashily and, between them, cause so much turbulence that a child also (independently) swimming in the lake is drowned.

What is the appropriate moral judgement on these teenagers? None, individually, has harmed the child: the minor turbulence caused by one alone would not have done so. Certainly, there is no (collective or individual) *intention* to harm. Nonetheless, it is as a result of the combination of their actions that the child drowns. Building on standard moral thinking, I suggest that whether we hold the teenagers responsible on a particular occasion is likely to depend on whether they could reasonably be expected to be aware of the danger.

If they are all unaware of the child's inability to swim well, of the likelihood that others besides themselves would be diving in, or of quite how much turbulence lots of teenagers jumping about in the water would create, it might be considered a tragic accident. If, however, they do not stop their behaviour as soon as the risk becomes apparent, or behave in exactly the same way on another occasion, with another child, then it seems that they are collectively responsible.<sup>33</sup>

The argument, "It wasn't just me – what I did wouldn't have done any harm if there hadn't been lots of others doing it as well" might have some appeal in saving the teenagers, *qua* individuals, from full responsibility. However, it seems plausible that it is sufficient to ascribe weak collective responsibility that the following is satisfied.

---

<sup>32</sup> May's example, recall, is of holiday makers who might, depending on their capacity to organise in time, be held collectively responsible for failure to save a drowning child. (1992, 110-6) However, my more positive sense of collective responsibility appears to be implicit in his account: he describes such "putative groups" as "most responsible for preventing harms that those groups *have done something to bring about*, or about which they have raised expectations". (1992, 118; emphasis added) But what would it mean for such a putative group to have *done something to bring about* a harm? There is no formalised or even informal structure such that the action bringing about the harm could count as collective in May's sense. The gap can be filled by the notion of weak collective responsibility.

<sup>33</sup> Indeed, to anticipate the discussion below, these conditions go further than those already familiar, for the individual case, from the moral foundations of tort law. There, for example, Perry suggests that "when common knowledge of the relevant causal regularities would lead *an agent of average mental capacities* to be aware of a sufficiently high level of risk of harm to other persons, taking account of both the probability and seriousness of the outcome, then the action should be treated for the purposes of reparation as faulty". (1992, 509-10; emphasis added)



*Two-part sufficient condition for weak collective responsibility:*

- the individuals acted in ways which, in aggregate, caused harm, and which they were aware (or could reasonably be expected to have foreseen) would, in aggregate, cause harm (although each only intentionally performed his own act); and
- they were all aware (or could reasonably be expected to have foreseen) that there were enough others similarly placed (and so similarly motivated to act) for the combined actions to bring about the harm.

It is important to distinguish this model from May's shared responsibility. Clearly, some of the same cases will be accommodated by each. However, there are two important distinctions, which the rest of this section will bring out.

Firstly, on May's model, each individual has a share of personal (*individual*) responsibility for the harm. My claim is that the group (in a weak sense) is *collectively* responsible. The implications are different. On May's account, presumably, individuals, *qua* individuals, have a duty not to do their "bit" of the harm (or to make reparation). With weak collective responsibility, there is a collective duty to end, or repair, the overall harm. Individual duties are acquired through that. (I will return to this in the next section.) The second distinction is that my model covers a wider range of cases.

Both points can be made through more detailed consideration of the ways in which a number of individuals might be said to "contribute to" a harmful result without being part of a collectivity.

Firstly, a number of individuals could each cause small harms. These sum together straightforwardly as a larger harm. Suppose, for example, the children in a village each steal an apple a day from a farmer's orchard. Each theft constitutes a small harm, and the sum of these small harms is a greater harm (the loss of many apples). The attribution of personal responsibility is unproblematic, and the *shared* responsibility of all the individuals is simply the aggregation of the

personal responsibility each has for the harm he has done.<sup>34</sup> Call such cases Aggregated Harms Cases.

Secondly, each individual could contribute to the overall harm, but without direct responsibility for any specific part of it. This could happen in two ways. Firstly, what I will dub Concerted Joint Action Cases, which correspond to May's understanding of joint action, above. His example, recall, is of thieves coordinating intentionally to steal buttons. The second possibility is that individuals contribute to a harm through actions which are not in themselves harmful (which may even be beneficial) and which are not intended to contribute to harm, or even to contribute to some common goal. Call such actions "harm ingredients" and such cases Aggregated Harm *Ingredients* Cases.<sup>35</sup> The teenagers in the lake present one example. At a wider level, Joel Feinberg uses the example of sulphur dioxide emissions which, in combination, render dangerous the percentage of the gas in the air.<sup>36</sup>

As we saw in Section II, May defends his model explicitly for Concerted Joint Action Cases, leaving it unclear whether he intends also to accommodate Aggregated Harm Ingredients Cases. He does indicate that shared responsibility would be appropriate in cases where a person acted 'with no knowledge or intent of contributing to a harm [but nonetheless] did perform a necessary part in a harm'.<sup>37</sup> However, it is arguable whether this is intended only to refer to a particular case of Concerted Joint Action, where the common end is some other goal of which harm is a (foreseen) side effect, or to capture the wider Aggregated Harm Ingredients Case, where there is no conscious pursuit of *any* shared end.

---

<sup>34</sup> Note that there may be cases in which the overall harm is greater than the sum of its parts - on top of being small harms in their own right, each individual action may (to anticipate the discussion below) be a harm ingredient contributing to some greater harm. Suppose the apple-stealing children take so many apples between them that the farmer's cider business fails. In such cases there is personal responsibility for each individual harm, but it doesn't capture the wider harm so well as weak collective responsibility.

<sup>35</sup> Although in Aggregated Harm Ingredients Cases the actions are likely to be similar ones, it is at least possible that they could be different actions which combine predictably but unintentionally to bring about some harm. For example, certain activities in a public place could (and could reasonably be expected to) combine to cause harm. Suppose one of the people-covered-in-paint-pretending-to-be-statues in Covent Garden were to use flammable paint and occupy a stand next to a flame juggler. Subject to similar criteria as for the teenagers in the lake case, the two could be collectively responsible for harm to passers by (and the statue-actor himself) from an explosion.

<sup>36</sup> Feinberg 1987, 228

<sup>37</sup> May 1992, 39

Let us assume that the shared responsibility model is intended to cover at least some Aggregated Harm Ingredients Cases. However, upon examination, May's account is rather less appealing in such cases than in Concerted Joint Action Cases. Given that there is no intention to contribute to a collective act (never mind an intentionally harmful one), the idea that responsibility *for harm* can be assigned to individuals, *qua* individuals, is less than compelling.

Recall the three possibilities outlined by May for allocation of personal responsibility: division of responsibility into personal responsibility for shares of the harm, no responsibility to each, or responsibility for the full harm to each.<sup>38</sup> As we saw in Section II, May himself argues that neither no nor full responsibility seem appropriate.<sup>39</sup> This leaves us with the first option, which is plausible in at least some Concerted Joint Action Cases, but less so in Aggregated Harm Ingredients Cases.<sup>40</sup>

How are individuals to be assigned "pieces" of the harm, according to the roles they played, when their actions, which were *not* in themselves harmful, did not constitute the performance of roles in some deliberate collective endeavour? Yet it seems equally arbitrary simply to take up the other option suggested by May, and divide up the (perhaps very significant) harm and say each of *n* individuals was responsible for 1/*n*th of it.

The point is precisely that the individual, *qua* individual, has not harmed. Nor has she intentionally contributed to a joint harm. Given this, it seems inappropriate to talk of each individual being assigned a share of *direct* personal responsibility, *at all*.<sup>41</sup>

Consider the alternative: collective responsibility. On May's account, recall, this is appropriate only if the group is structured in such a way that the harmful action could not have occurred if the individuals were acting outside it. Both Aggregated Harm Ingredients and Concerted Joint Action Cases have something in common with this. While, in these two scenarios, the *action* is not collective in

---

<sup>38</sup> Ibid., 42

<sup>39</sup> Ibid., 39

<sup>40</sup> Recall May's appeal to Prosser and Keaton (May 1992, 41)

<sup>41</sup> To use the distinction to which May himself appeals in the case of collective *inaction*, it would generally seem inappropriate to blame them, as individuals, for the harm. Moreover, what they should feel is not the guilt of one who has intentionally harmed, but rather shame that a pattern of actions of which they were wittingly a part, has resulted in this harm. (1992, 38)

the strong sense of irreducibility to individual actions, the *result* (harm) could not have occurred were not those individuals situated, in relation to one another, in such a way that their pursuit of individual goals would have a certain predictable aggregative impact. It is only because of the way the individuals are *grouped*, in this very weak sense, that harm is done. Thus, it makes sense to assign moral responsibility collectively, albeit in a similarly weak sense, rather than attempt to divvy up the harm into individual responsibility “shares”.

I contend, then, that while at least some Concerted Joint Action Cases do plausibly involve shared personal responsibility, they can *also* be cases of collective responsibility.<sup>42</sup> (Indeed, in at least some such cases it seems that there may be a stronger sense of collective responsibility – the collective responsibility of an informal but intentional group such as the mob, to which May has earlier attributed joint *action*.<sup>43</sup>) Moreover, weak collective responsibility is more appropriate than May’s notion of shared responsibility in Aggregated Harm Ingredients Cases.

Turning to the second promised distinction, there are important cases apparently explicitly excluded by May but covered by weak collective responsibility. Upon examination, Aggregated Harm Ingredients Cases fall into two categories. One includes those meeting May’s criterion for joint action: each individual’s action is necessary (but not sufficient) to bring about the harm.<sup>44</sup> The other includes those cases where the number of contributory actions is above the threshold to trigger the harm. Thus, although the combination of actions brings about harm, each individual’s contribution is not necessary for that harm. Call these Overdetermination Cases.

My model, leaving it open whether each individual’s action is necessary, allows for Overdetermination Cases. This is crucial, given that it is frequently in such scenarios (for example, the aggregation of individual decisions to drive cars rather than use public transport, or to fly abroad for a holiday, resulting in significant environmental damage) that ‘an important outcome [is] the product of

---

<sup>42</sup> The attribution of weak collective responsibility (rather than personal responsibility) becomes still more compelling when the harm is not the common goal to which each individual consciously contributes, but an unintended side effect.

<sup>43</sup> May 1987, 31-41

<sup>44</sup> May 1992, 39

a large number of acts performed by *many* different people, few if any of whom actually intend to produce [it]’.<sup>45</sup>

#### IV. Intentionality and overdetermination

Having laid out the weak collective responsibility model, I will defend it by responding to two potential objections. The next section will expand on the model’s implications in terms of collective and individual duties, and in so doing reject an additional objection.

The first objection appeals to the popular idea that moral responsibility requires intentionality. There is, it points out, no intention to harm in cases where harm results from the aggregation of individual acts, and so there cannot be moral responsibility.

However, it is questionable to what extent responsibility should rest on intention. Even at the individual level, it is often sufficient for the ascription of moral responsibility that an individual could reasonably have been expected to know that what she was doing would cause harm.<sup>46</sup> While there is considerable intuitive appeal in the idea that, on a one off occasion, not knowing might be an excuse (“I didn’t know that pressing this switch would cause my neighbour’s house to blow up”), this loses its force if the individual continues to behave in the same way once the connection is made clear to her. (“Yes, I have read all the statistics on the effects of alcohol on ability to control a vehicle, and, yes, I was drink driving, but I didn’t *intend* to run over that man, so I’m not morally

---

<sup>45</sup> Scheffler 2001, 43 [emphasis added] Note, however, that I have not rejected May’s account altogether. Nor do I need to. It is both plausible and probable that there will also be shared personal responsibility in (at least some) cases of weak collective responsibility. Recall that May himself expects overlap between shared responsibility and his stronger collective responsibility. (1992, 107) Indeed, while my account is primarily forward-looking – *accurately* assigning responsibility with a view to attributing duties so as to bring about the morally optimum end to a harm – often, in practice, we need ways of assigning responsibility as fairly as possibly between *individuals*, in a backwards-looking sense, so as to hand out fines or punishments.

<sup>46</sup> This might at first sight appear to run contrary to the so-called doctrine of double effect (DDE): ‘[T]he thesis that it is sometimes permissible to bring about by oblique intention what one may not directly intend.’ (Foot 1978b, 20. The relevant chapter is reproduced from *The Oxford Review* 5, 1967) However the DDE does not rule out the possibility that it *is* on occasion impermissible to bring about an effect even obliquely. Moreover, it has been called into doubt by philosophers including Foot herself. (See 1978b, 19-31)

responsible.”)<sup>47</sup> The moral force, in such scenarios, derives from awareness of the potential consequences; intention to do a harmful act is not required.

This is reflected in tort law, which, through the notion of recklessness, incorporates the idea that individuals should be held responsible (at least to the extent of a duty of reparation) not only for what they intend to do but also for consequences they foresee as a side effect of their action.<sup>48</sup>

The second objection goes as follows. I have made it clear that weak collective responsibility, unlike May’s shared responsibility, accommodates Overdetermination Cases. But if what an individual did made no difference, why should she be one of those held collectively responsible? The answer, I suggest, is, precisely because *any one* of the individuals was similarly unnecessary – it would be arbitrary (and unfair) to identify the harming group as all acting in that way less one specific individual. (This idea receives further support from the notion of morality as a collective project, borrowed from Liam Murphy and discussed in the next section.)

That said, some question remains as to what the appropriate cut off point is for inclusion as one of those collectively responsible in such cases. Otherwise, people could become collectively responsible for various harms merely by existing.<sup>49</sup> This is counterintuitive. For example, everyone has some carbon footprint, but that does not necessarily render everyone including (say) a South American rainforest tribe collectively responsible for environmental harm. A plausible working suggestion would be that a person becomes one of the group collectively responsible for harm once her contribution exceeds the amount such that, were everyone contributing only to that level, there would be no harm.<sup>50</sup>

---

<sup>47</sup> Of course, there might be mitigating circumstances – preventing worse harm, for example – but I will save such discussions for Chapter Five.

<sup>48</sup> In fact, in also incorporating *negligence*, tort law goes further: it rests liability not what the relevant individual foresaw at the time, but on what a reasonable person would have foreseen in the circumstances.

<sup>49</sup> It is, of course, possible to imagine some dire situation in the environmental harm case – perhaps an overpopulated future – in which this was the case. In such a scenario, no doubt the cost of not harming (see Chapter Five, Section V) could be applied to as a way to render the harm morally permissible. But this is not the case at present.

<sup>50</sup> As Singer points out, in terms of per capita emissions and even stabilising climate change, residents of the US, Japan and Western nations are already several times above this level. (2002, 35) However, applying this condition is not always straightforward. In particular, problematic scenarios arise in the environmental case, because of population growth. The more people there are, the lower the level of emissions will be at which one becomes part of a collectively harming group. But this allows for the possibility that Z, who has emitted at a constant, very low level and

Having thus defended the notion of weak collective responsibility, it remains to consider the correlative duties, both collective and individual. A further objection will be addressed within this discussion.

## V. Collective and individual duties

What duties are implied by weak collective responsibility for harm? When an individual has harmed, there is generally a duty (assuming the harm is morally impermissible) not to (continue to) do the harmful act. (And, quite often, to make reparation for harm already caused.)

With weak collective responsibility, the primary corresponding duty is naturally a collective one: to do something about it together. In practice, this is likely to require collective decision making. Especially in larger groups, while the aggregation of individual decisions to cease acting so as to contribute to harm could sometimes end that harm, motivational problems render it implausible that this would in fact happen.<sup>51</sup> Moreover, coordination problems will frequently render this an inefficient mechanism, sometimes an ineffective or even counterproductive one.<sup>52</sup>

In smaller groups, such collective decision making might be more informal. (In the lake case, the teenagers might decide among themselves to take turns to swim.<sup>53</sup>) However, larger harming groups are likely to require a more formal structure or institution before they can decide, in any meaningfully collective sense, how to end the harm. Thus, where there is not already such an institution in place (which is likely to be the case where there is *weak*, rather than the stronger corporate collective responsibility) the collective responsibility will

---

had no children, becomes one of those collectively responsible when he wasn't before because A to Y have had several children each, increasing the population. I cannot respond to this fully here, but would certainly suggest that one's own carbon footprint include a factor for the number of children one has had.

<sup>51</sup> Shue, in the context of duties to protect basic rights to subsistence, highlights the likelihood of cases 'in which individual restraint would be too much to ask'. (1980, 59)

<sup>52</sup> This claim will be defended in detail in Chapter Four. In theory, if each individual acted, independently, in a certain way, the harm would cease. However, in practice such uncoordinated action is likely to be an inefficient (or even, in such circumstances as are discussed later in this section, a morally suboptimal) way of ending the harm.

<sup>53</sup> That said, with a group of changing membership, enforceable restrictions are likely to be necessary. (In practice, this is what happens in public swimming pools.)

most often give rise in the first instance to a collective duty to establish such an institution.<sup>54</sup>

A number of individual duties, depending on the circumstances, correspond to this collective duty. These, I suggest, fall into two categories: duties to play one's part in bringing about that collective end to the harm, and supplementary individual duties.

The former can, again, take two forms. Firstly, duties to fulfil the individual requirements of a collectively pursued solution to the harm. Recall that the collective duty, put most generally, is a duty to change the situation so that the harm is not caused (or reparation is made). Assuming that some way has been established of making collective decisions, conclusions regarding appropriate action would be handed down as restrictions on, or requirements for, action by individuals. In this way, assuming fairly-established decision-making structures and fairly-allocated duties, members of the harming group would acquire morally-grounded individual duties.<sup>55</sup>

Secondly, even in the absence of the necessary collective decision-making structure or institution (and so of clearly defined restrictions on individual behaviour), individuals, by being part of a harming *group*, acquire certain responsibilities which, I suggest, they can no longer evade simply by "opting out".

Reconsider the lake case, with two supplementary assumptions: this is an Overdetermination Case, so the removal of any one of the teenagers would not save the child, and the conditions for weak collective responsibility are met. Each teenager could reasonably be expected to foresee that the combination of his action with others would cause harm, and knows that there are others sufficiently motivated to act, for the harm to be the likely result. Now suppose that one of the teenagers *does* foresee the harm – he realises what is going on. What are his options?

---

<sup>54</sup> Such a duty is not a new idea – merely an often neglected one. Rawls posits, as part of the (natural) duty of justice, not only a responsibility to obey and support just institutions, but also a duty "to further just arrangements not yet established, at least when this can be done without too much cost to ourselves". (1971, 99) Shue, in his account of subsistence rights as basic rights, advocates a general duty to design institutions "that avoid the creation of strong incentives to violate [the] duty [to avoid depriving others of their subsistence rights]". (1980, 60)

<sup>55</sup> These questions of fairness constitute a huge topic in themselves, but one I will reserve for future research.



The teenager could simply remove himself from the harming set (in this case, climb out of the lake) without saying anything. This seems inadequate. There is more at stake, morally, for each individual, than just his own contribution.<sup>56</sup> Rather, it seems, he has a responsibility, as part of the harming set, to make an effort to promote an end to the *collective* harm. In this case, he would most probably fulfil that duty by alerting the others and promoting a way of avoiding the problem, such as a jumping rota. At a wider level, promotion is likely to have to take the form of political advocacy of certain collective decision-making institutions.

Thus, I suggest that individuals who belong to groups collectively morally responsible (even in the weak sense) for unacceptable harms have a duty actively to promote the establishment of such collective decision-making institutions as would be necessary to bring about an end to the harm.<sup>57</sup>

Arguing parallel to May, this duty could be said to fall more heavily on certain individuals who are in a position to be heard.<sup>58</sup> Pre-empting Chapter Five to assume that the environmental damage case *does* count as unacceptable harm to future generations and the global poor, the relevant individuals would be those wielding global political influence. These would include: prominent political leaders, those with economic or financial clout (such as World Bank officials), those with authority in the most polluting industries (such as oil company management), and even (arguably) those with religious authority.<sup>59</sup>

---

<sup>56</sup> Unlike Pogge in the global justice case (2002, 66), I do not think my becoming the environmental equivalent of a hermit would constitute the satisfaction of all the duties correlated to collective responsibility for environmental harm. Rather, I would have first to at least try for some time (a month? a year?) to promote the establishment of institutions to address environmental damage collectively.

<sup>57</sup> Pogge, in the rather different context of responsibilities of members of existing unjust institutions, posits a *negative* duty on all human agents 'not to co-operate in upholding [coercive institutions under which the access of some members to the objects of basic human rights is insecure] unless they compensate for their co-operation by protecting its victims or by working for its reform'. (2002, 67) However, in both his case and mine, worries are raised about overdemandingness. How far does an individual have to go in promoting a collective end to the harm, before he can legitimately opt out? In this particular case, to anticipate the discussion below, if his attempts to change the behaviour of the rest were to fail (or even look likely to fail) then, rather than opt out, he should attempt to rescue the child by removing it from the water. More generally, it seems that there are limits to the sacrifices one individual can be required to make.

<sup>58</sup> May, recall, appeals to such considerations in attributing personal responsibility for harm in collective inaction cases. (May 1992, 112-6. See Footnote 31)

<sup>59</sup> The Pope is an obvious example, although the Catholic Church's policy on contraception, in adding to the population problem, is in practice making matters worse.

One final point before moving on to supplementary individual duties. This account draws a clear distinction between members, or ex-members, of harming groups, and bystanders, which in itself calls for some further explanation. Does not the bystander, in such a scenario as that of the teenagers in the lake, *also* have an obligation to try to prevent the harm? Of course she does. However, hers are *positive duties* to help the child. The (ex-)member also has those duties, but as one of the harming group he has *also* acquired negatively derived duties, including the duty to promote, if at all practicable, a *collective* end to that harm, of which opting out has not absolved him.

This leaves the twofold issue of what I have called supplementary individual duties: duties that might be incurred alongside the promotion process, *if* it has at least a reasonable chance of success, and those incurred in the place of promotion duties if there is, in practice, little chance that collective decision-making institutions will successfully be established.

In either case, one or both of two further duties need to be considered. Firstly, the duty to mimic (or even exceed) the duties which would be assigned by a collective decision-making body. However, at least in the first scenario (promotion of a hopefully achievable decision-making institution), the duty to promote the institution will almost invariably take priority. As noted above, and as will be argued in detail in Chapter Four, coordination and motivation problems make it frequently implausible, or even impossible, to achieve a common or collective aim through the aggregated behaviour of uncoordinated individuals.

Even supposing, however, either that they can be fulfilled simultaneously with promoting the collective institution, or that there is no realistic chance of bringing about the institution, it can be problematic to assign these individual “mimicking” duties. What would such action amount to? Most probably, cutting down on or stopping one’s own individual contributing action. That is, in the lake case, getting out of the water unless there are few enough others for it to be safe, or, in the environmental one, reducing one’s own carbon footprint, not having more than two children, and so on.

Such examples bring out an important difference between Aggregated Harms Cases and Aggregated Harm Ingredients Cases. There seems to be a much stronger moral incentive to refrain from the individual act if the overall harm is

an aggregation of small *harms*, and so the individual refraining could prevent one particular sub-harm (as with the animal rights campaigner who goes vegetarian, reasoning that she is at least saving one cow), than if the acts are not themselves harmful.

In these latter, Aggregated Harm Ingredients Cases, there is likely to be a threshold effect. A certain number of people need to change their behaviour before any benefit will be felt. (If I reduce my carbon emissions to the level that would be required of me by some hypothetical global climate change body, but no-one else does, this will make absolutely no difference to environmental damage.) But in that case, so long as it is improbable that anyone else *will* change their behaviour – or that this one sacrifice will motivate more – it becomes a symbolic act and, arguably, nothing more than moral “showing off”.<sup>60</sup>

The second potential duty is that of attempting to mitigate the damage, oneself or in combination with a likeminded subset. For an individual teenager in the lake case, this could involve attempting to rescue the child. In the environmental case, it might take the form of regeneration efforts, or of providing support to those who live in areas already affected by climate change. In practice, this may well run alongside the primary duty of campaigning for the relevant institutions to be set up. In the absence of any realistic hope of establishing institutions to prevent the collective harm, it might be the only option.

However, there is another possible scenario to be addressed, in which this last duty could overlap with the earlier duty to bring about an end to the collective harm, and which gives rise to the supplementary objection left open at the end of the last section.

The scenario concerns those Aggregated Harm Ingredients Cases which are not characterised by overdetermination: where each individual’s contribution is necessary for the harm, an individual has the option of opting out *and so ending the collective harm*. This explains why it is more plausible in such cases than in Overdetermination Cases to assign at least some personal responsibility. Could it, further, resuscitate the question of whether the idea of weak collective

---

<sup>60</sup> That said, it is worth noting that such a sacrifice, by enough people, could plausibly be helpful in promoting a collective institution to organise a wider effort. On the flip side, it is very possible that those promoting such an institution without making such sacrifices themselves, could be dismissed by the public as hypocrites and so undermine their promotion attempts.

responsibility is significantly distinct from the shared responsibility account offered by May? Does my model *only* add value in Overdetermination Cases?

I think not. The arguments of Section III, which distinguished between weak collective and shared responsibility for harm, still hold. Moreover, such solutions (the opting out of one individual to end the harm) are rarely optimal, because they are generally unfair. The relevant individual takes on the full burden, regardless of how much more others have contributed to the harm, and of the possibly greater harmful side effects of ceasing her contributory act than those of others – or of reducing them across the board.<sup>61</sup>

In parallel with Murphy's discussion of principles of beneficence, one might make this point by appealing to the idea of morality as a collective project. He cites the 'natural thought... that it is objectionable to expect agents to take up the slack caused by the non-compliance of others. We should do our fair share, which can amount to a great sacrifice in certain circumstances; what we cannot be required to do is other people's share as well as our own.'<sup>62</sup>

However, there is a relevant difference between the cases of beneficence considered by Murphy and those addressed here, of collective *harm*. While his "natural thought" supports the idea that the best outcome in weak collective harm scenarios is a collective end to the harm, it would be too strong to say that no individual can, in such cases, ever "be required" (morally speaking) to take on the shares of non-compliers.

Compare the following two cases. In the first, I am one of a number of bystanders on the scene of a minor car crash. This accident was the result of natural causes. No-one is hurt, but the car has become stuck at the side of the road. If all the bystanders present helped, it would require only a few minutes of our time to free the car. However, many don't help. If I stayed to move it on my own, the time taken would be longer, and the cost to me greater (say, I would miss a significant business meeting). According to Murphy's plausible Compliance Condition, my duty according to a principle of beneficence should not extend beyond doing what would have been my share in full compliance.<sup>63</sup>

---

<sup>61</sup> Feinberg makes this point. (1987, 227-32)

<sup>62</sup> Murphy 1993, 278

<sup>63</sup> '[A] principle of beneficence should not increase its demands on agents as expected compliance with the principle by other agents decreases.' (Murphy 1993, 278)

Consider now the alternative scenario in which I am one of several individuals collectively responsible (in the weak sense) for this harm. (Suppose it can be safe to overtake on that road, but only if no-one else does, and we all overtook knowing this and that many other drivers were likely to think in the same way.) Now, if the other drivers leave without helping, while they have undoubtedly acted wrongly to the victim (and unfairly to me), it is no longer so plausible for me to say: "I've done my share of the clearing up. I'm going now." It is at least arguable that, having contributed to the harm, I should stay until it is cleared up.<sup>64</sup>

My model can explain this intuition. It can both accommodate a duty to opt out of the harming set, if there is no alternative way of ending the harm, and acknowledge that this is not generally the morally optimal solution. All individuals in the collectively harming group have a duty to endeavour to bring about a collective end to the harm. In the absence of cooperation from others, if an individual can bring about the collective solution single-handedly, i.e. by opting out, then, often, he should. ("Often" because the potential moral "excuses" to be addressed in a later section may come into play.) But he is only obliged to do so because of the failure of others to act as *they* should – a line of thought which lends further support to the idea (to be examined in Chapter Five), that certain collective decisions to end collective harm, can legitimately be enforced.

## Conclusion

This chapter has argued that moral philosophy needs to accommodate the concept of weak collective responsibility for the harm which results from the predictable aggregation of individual acts. I have argued, further, that this collective responsibility gives rise to a collective duty to (to put it most broadly) do something to resolve the situation. Correlated to this are individual duties: as

---

<sup>64</sup> This line of thought brings out, again, the difference between bystanders and ex-members, as discussed above. The reason, despite this distinction, that the demands on the two might sometimes, in practice, be very similar, is that sometimes positive claims of beneficence are so strong as to demand a significant sacrifice even a bystander. Murphy himself allows that the Compliance Condition may not hold in cases of special obligation, in which he includes rescue cases. (1993, 290-2)

part of the collective endeavour (promoting the establishment of collective decision-making structures, then abiding by their decisions) and, potentially, either alongside it or in its place if the lack of cooperation by others renders it impossible.

This has important implications. At the close of the next chapter, I will use the notion of collective responsibility for harm to identify what I will call *morally salient potential collectivities* – groups of individuals who do not yet constitute a collectivity but ought, on moral grounds, to count (politically) as though they did. In Chapter Five, I will argue that there is weak collective responsibility for the harm done through damage to the environment, so we, or at least most of us, in this generation, acquire certain corresponding duties. (In other words, we constitute just such a morally salient potential collectivity.)

Before closing, there are two final points to be made, to which I will also return. Firstly, I have assumed, throughout this chapter, that is, that it is philosophically possible to flesh out what we mean by doing harm, in a morally relevant sense, to other human beings. In Chapter Three, I will use a capabilities model of human flourishing to justify this assumption.<sup>65</sup>

Secondly, even with a clear understanding of morally significant harm, it is not invariably considered morally *impermissible*, at the individual level, to do such harm to another person. There are various moral “excuses” to which we might appeal. These excuses have varying degrees of plausibility, but do require consideration. I will return to them, in Chapter Five, to contend not only that we are collectively responsible for environmental harm, but that this harm is morally impermissible.

---

<sup>65</sup> I will claim that undermining one of the central capabilities on something like Nussbaum’s list counts as morally significant harm. (Nussbaum 2000b, 78-80) I do not claim, however, that an effect *only* counts as harmful if it undermines one of these central capabilities.

## Chapter Two

### Against the Intentionalist: Collectivities v. Social Groups<sup>1</sup>

Under what circumstances do a set of individuals constitute the kind of collectivity with which political philosophers should be concerned?

Historically, political philosophy has focused almost exclusively on the single society – and, as Scheffler notes, tacitly assumed ‘a one-to-one correspondence between individual societies and sovereign states’<sup>2</sup>. This has at once been supported by and posed problems for the prominent intentionalist tradition in the philosophy of social science. According to the intentionalists (broadly speaking), in order to constitute a collectivity or social group, a group of individuals must all consider themselves to be part of such a group.

There are two reasons to reassess the philosophy of social science behind the understanding of collectivities standardly accepted in political philosophy. On the one hand, there are practical incentives. Humans no longer live in neatly separated, self-sufficient units. International communication and travel have facilitated an ever-expanding network of economic and legal ties. Moreover, between us, we have an impact on the environment which affects us all. It is time to reconsider the question: within which sets of relations do we acquire special responsibilities as co-members of some collectivity?

On the other hand, as this chapter will argue, the dominant intentionalist model, Gilbert’s plural subject account of collectivities, is inadequate to account for even the collectivity on which political philosophy *does* focus: the nation state. Gilbert, treating collectivities and social groups as the same, and defining them via acknowledgement of shared goals, ends up unable to accommodate certain members of the nation-as-collectivity, such as babies and the mentally handicapped.

This chapter will engage in particular with three philosophers of social science. Like non-intentionalists Paul Sheehy and Keith Graham, I will argue that Gilbert’s plural subject account specifies conditions which are unnecessary for collectivityhood. However, I will also contend, against all three thinkers, that

---

<sup>1</sup> Much of the material for this chapter is drawn from my MPhil thesis (Cripps 2005). However, I will only refer specifically to this thesis when I am alerting the reader to arguments offered there which it is unnecessary, for the purposes of my wider project, to reproduce here.

<sup>2</sup> Scheffler 2001, 32

social groups are only a subset of the wider set of (politically and explanatorily interesting) collectivities. In so doing, I will argue that the kinds of collectivities with which political philosophy should be concerned include much wider groupings than the nation state.

The structure is as follows. Section I will lay out the non-intentionalist model to be defended: a set of individuals constitute a collectivity if and only if they are mutually dependent through some common or shared interest, purpose or goal, on a certain broad understanding of common interest, purpose or goal.

Section II will summarise Gilbert's plural subject account. Section III will argue that her argument fails to accommodate certain key collectivities and, in the process of trying to do so, becomes circular. Sections IV to VII will show how the non-intentionalist model offered in Section I can accommodate these troublesome cases without circularity. In order to demonstrate this, I will elaborate on the distinction between social and collective properties, and between collectivities, of which social groups are a subset, and *aggregates* (Sections V and VI).

Having thus defended the non-intentionalist model as able to accommodate the collectivities already key in our social and political thought, I will defend its inclusion of more controversial cases. Section VIII will appeal to some controversial, largely non-intentional examples. Section IX will argue that all the collectivities on my account satisfy three conditions for collectivityhood laid out by Graham: the significance of the actions of the individual members cannot be adequately captured without referring to the collective body as part of which individuals are acting, what that collective body does is distinct from what its individual members do, and the survival of the collectivity is '*relatively indifferent to the persistence of the particular individuals which compose it at any particular moment*'.<sup>3</sup>

Section X will respond to two obvious objections: that the model is too demanding, and that it is too inclusive. Finally, Section XI will link this chapter with the previous one, introducing the idea of a morally salient potential collectivity, as also meriting the attention of political philosophers, and potentially grounding special obligations.

---

<sup>3</sup> Graham 2002, 68-9 [emphasis in original]



## I. The non-intentionalist model

This chapter rejects the intentionalist claim that, in order for a set of individuals to constitute a collectivity, it is necessary and sufficient that each regards himself as part of a group constituted by him and the others. Instead, the following model will be defended.

A set of individuals constitutes a collectivity if and only if those individuals are mutually dependent through some common or shared purpose, goal, or all-things-considered (ATC) interest, whether or not they acknowledge it themselves.<sup>4</sup>

Alternatively, this can be expressed as:

A collectivity is constituted by a set of individuals of whom at least one collective property in the strong sense is true.

Where:

Property *c* is a collective property in the strong sense if and only if the applicability of *c* to a set of individuals entails that those individuals are mutually dependent through some common or shared goal, purpose or ATC interest.

The “mutual dependence” clause requires that the individuals have goals or purposes or ATC interests that it only makes sense to consider as being pursued together. This can be understood in two ways: they could have a genuinely shared goal, as with a group of friends setting out to climb Monte Rosa *together*; or they could have individual (selfish) goals, ATC interests or purposes which require co-operation by all for their achievement. Consider, for example, a group

---

<sup>4</sup> Chapter Three will explain the limited range of cases in which an individual can be said to have an ATC interest which she does not regard as a goal or purpose. The idea is that enjoyment of certain central capabilities constitutes such a fundamental human interest, being necessary for a fully flourishing life, that it is not in a person’s ATC interest to give up one of these capabilities, or expose it to serious risk.

of strangers washed up on a desert island, none of whom care about each other's welfare, but who need to co-operate in order to build a shelter to survive. (Something not unlike the start of the television series, *Lost*.)

Note that this allows for both goals, purposes or interests which, by their nature, can only be pursued together (for example, the playing of a symphony by an orchestra), *and* goals, purposes or interests which in the circumstances can only (can only efficiently) be pursued together. An example of the latter would be the keeping clean of a vast mansion.

Importantly, the group must include all and only those who are thus mutually dependent. It would be arbitrary to include some but not all of the friends on Monte Rosa, or all of the London Symphony Orchestra save one flautist.

Given such a broad account, there are many distinctions to be drawn. Collectivities can be new or pre-existing. They can be voluntary or involuntary. They can be small or large. They can be passive or active. They can be *ad hoc* (constituted by individuals who are mutually dependent only through a limited, short-term goal, purpose or interest) or lasting. They can also overlap categories, so that (for example) for some members a group is voluntary, and for others involuntary. Crucially, they can be intentional, partially intentional (as with the family or nation state) *or*, most controversially, entirely unintentional.

Some but not all of these collectivities are also social groups. Here, this thesis breaks with both Gilbert and her non-intentionalist critics, Sheehy and Graham. Since my primary aim is to expand the account of *collectivities* with which political philosophy should be concerned, a thorough examination of the notion of the social will not be necessary. Instead, I refer the interested reader to my earlier treatment of this topic.<sup>5</sup> However, the following rough definitions, to which I will return in Section VI, can be used to clarify the distinction and accommodate certain key cases whilst avoiding circularity problems.

A set of individuals constitute a social group if and only if they constitute a collectivity and: either, it is because of at least one social property true of the individuals that they constitute a set of which a collective property is true, or it is an integral part of the common goal, interest or purpose

---

<sup>5</sup> Cripps 2005

through which the individuals are mutually dependent that at least one social property becomes or remains true of someone, some set of individuals, or something.

Where:

It is a necessary condition of a property's being a social property that it has a significance over and above the material or physical properties with which it coincides, and has this because of the recognition by a certain group of individuals that it will have it.

Sharing a social property is neither necessary nor sufficient for collectivityhood.

Rather:

A set of individuals constitute an aggregate if and only if some property is true of all those individuals, but they do not constitute a collectivity.

A set of individuals constitute a social aggregate if and only if some social property is true of all those individuals, but they do not constitute a collectivity.

The following examples will illustrate the range of this account. They can also be used to show: firstly, the non-intentionalist model's ability to accommodate the intuitive, small-scale collectivities to which intentionalist models such as Gilbert's are also well suited, secondly, its ability to account for those bigger groupings of individuals which intuitively (and politically) count as collectivities but with which (as we shall see) Gilbert's account struggles, and finally, some of the more controversial groupings which it also counts as collectivities.

An example of a new, small, voluntary, intentional collectivity is the metaphysics reading group formed by six MPhil students from UCL and Birkbeck, with the intention of reading and discussing metaphysics papers at meetings each week. A college drama society is a pre-existing but voluntary, intentional collectivity. Fathers for Justice is an active subset of the wider and otherwise passive collectivity of divorced fathers.

A nation state such as Spain is a large, mostly involuntary, mostly intentional collectivity. A family consisting of parents and two children, one of whom is a rebellious teenager and the other of whom is a six month-old baby, is also a collectivity, albeit involuntary and unintentional for some of the members.

The final set of four examples takes us into more controversial territory. Firstly, consider the owners of weekend homes set round a green in a Gloucestershire village. By picking a commuter village, without any history of community feeling, we can assume that the owners do not consider themselves part of a group, rendering the collectivity non-intentional in a broad sense. Each wants the green to be pleasant, but only because this meets his or her individual requirements. However, they can only keep it so by co-operating. Each of them (say, by playing loud music or dropping litter) could undermine the pleasant atmosphere.

Secondly, on a variant of an example by Sheehy, four egoist prisoners break out, independently, from a jail surrounded by water.<sup>6</sup> They get into a boat with four oars and row away. The boat's size is such that no one of them, alone, would be able to row it, although each prisoner, being delusional, believes himself to be possessed of such superhuman strength that he could do so. Each thinks only of himself, "I", as escaping, but as a matter of fact the prisoners are dependent on one another for the achievement of each one's individual goal (escape). On my objective account, the prisoners constitute an *ad hoc* collectivity.

Thirdly, suppose four young men and women, members of deeply opposed gangs in their home city, find themselves cast up on an isolated island. As with the *Lost* example, co-operation to find food, build shelter, live peacefully and so on, is in each individual's interest. However, such is their loathing of each other, that they would vehemently deny any such idea, and in fact spend their time on the island attempting to kill each other. They constitute an involuntary, unintentional collectivity. (Call this the Rival Gangs Case.)

Lastly, and most controversially of all, Searle's example of a pack of hyenas. Says Graham:

A pack of hyenas can coordinate in very subtle ways in collectively stalking and killing a lion. Though it may be presumed that they have intentions in doing so, they do not

---

<sup>6</sup> Sheehy 2001, 128-130 and 2002, 384-5.

have the linguistic apparatus necessary for being aware of the intentions they have, and so *a fortiori* they are not aware of any such intentions being collective ones.<sup>7</sup>

I will return to these examples in later sections. Now, having laid out the model to be defended, I will now turn to the intentionalist account, rebuttal of which constitutes the negative component of that defence.

## II. Gilbert's intentionalist model

The intentionalist claim, put broadly, can be expressed as follows:

### *Broad Intentionalist Thesis*

In order for a set of individuals to constitute a collectivity or social group, it is necessary and sufficient that each regards himself as part of a group which is constituted by him and the others in the set.

This chapter will deny this claim, but will also reject the following more detailed version.<sup>8</sup>

According to Gilbert, social groups are plural subjects.<sup>9</sup> Plural subjects, in turn, are defined in terms of joint commitment.

A and B (and...) (or those with feature F) constitute a *plural subject* (by definition) if and only if they are jointly committed to doing something as a body – in a broad sense of ‘do’.<sup>10</sup>

Gilbert lays down three conditions for the creation of joint commitment. Firstly, it must involve all parties (all future members of the plural subject). Secondly,

---

<sup>7</sup> Graham 2002, 71-2, citing John Searle: *The Construction of Social Reality* (The Free Press, 1995)

<sup>8</sup> This project engages with Gilbert's plural subject account, because it is so thorough and has been so influential. However, Gilbert acknowledges her own debt to Georg Simmel (Gilbert 1989, 205, citing Simmel (1908): “How is society possible?” as reproduced in D N Levine ed. *Georg Simmel: On Individuality and Society Forms*, University of Chicago Press, 1971), and other theorists have offered intentionalist arguments. See, for example, Tuomela 2002 for a detailed “we”-based account of social groups.

<sup>9</sup> ‘In a nutshell, the proposal to be discussed is this: a social group is a plural subject.’ (Gilbert 2006, 165). She lays out her account in rather more complex detail in *On Social Facts* (1989), and I have previously defended the narrow intentionalist thesis as a plausible reading of the account given there. (Cripps 2005, 15-20)

<sup>10</sup> Gilbert 2006, 144-5

each party must, through their behaviour, make an 'expression of readiness for *joint commitment*: each understands what a joint commitment is, and expresses all that is needed on his or her part to bring such a commitment into being, namely, readiness to be jointly committed'.<sup>11</sup> Finally, there must be common knowledge among the parties that this is the case.

[I]f some fact is common knowledge between A and B (or among members of population P, described by reference to some common attribute), that fact is entirely out in the open between (or among) them, and, at some level, all are aware that this is so.<sup>12</sup>

Gilbert stresses that these expressions of joint readiness 'include but are not limited to what would count as everyday agreements'.<sup>13</sup> Joint readiness may be also expressed in action. (Gilbert's example is a number of people running towards the victims of an accident.) Its expression may even emerge gradually (as when two workers form the habit of meeting after work for a chat every evening).<sup>14</sup> Nor need the joint commitment be to some goal, narrowly defined. Rather, she notes, a set of individuals can constitute a plural subject by virtue of 'a single collective goal, belief, value or whatever'.<sup>15</sup>

She expands her account to accommodate large populations, where two conditions must be satisfied.

First, all members of the population must have expressed their readiness to participate in the relevant joint commitment with all other members of the population. Second, this must be population common knowledge.<sup>16</sup>

The former requires: that members of the population have a conception of that population, that they are all ready to enter into the joint commitment with all the others in that population, and that they express that readiness to the others. This, Gilbert argues, need not conflict with the anonymity of such situations. To use her example, if the members of a crowd shout their agreement to storm a corn-

---

<sup>11</sup> Ibid., 138 [her emphasis]

<sup>12</sup> Ibid., 138-9

<sup>13</sup> Ibid., 139. This, importantly, distinguishes the theory of political obligation defended in her latest work from actual contract theory.

<sup>14</sup> Ibid., 139-40

<sup>15</sup> Ibid., 167

<sup>16</sup> Ibid., 175

dealer's residence 'it is... natural to suggest that the expressions of readiness relate to those who are in the crowd – *whoever exactly they are*. In other words, each is ready to commit with those who fit a certain description – *being part of the crowd*.'<sup>17</sup>

Similarly, population common knowledge is 'common knowledge between people considered by those involved as members of a population individuated by a certain general description'. Thus, to offer a variant on her example, it might be out in the open that everyone living east of the mountain worships the Sun God, and everyone east of the mountain might be aware of that. 'Then – by definition – the existence of this situation is population common knowledge between those who live east of the mountain.'<sup>18</sup>

This model can be summarised as the following:

*Narrow Intentionalist Thesis*

a collectivity is a set or aggregate of individuals with (a) common purpose(s) or goal(s), all of whom are aware of that (those) goal(s) or purpose(s), aware that they are shared, aware of the awareness of other members of this, aware of each member's joint readiness to commit as a body to the joint goal(s) or purpose(s), aware of each member's awareness of *this* (and so on).

Lest this should seem too narrow to accommodate Gilbert's insistence that a set of individuals 'can constitute a plural subject by having a single joint commitment – hence a single collective goal, *belief, value or whatever*',<sup>19</sup> it is worth spelling out my understanding of "purpose". It includes commitment to beliefs or values. The idea is that shared *commitment* to such beliefs (or values) will have implications for the way one lives one's life: it will bring with it the intention to act in a certain way, in accordance with those beliefs or to maximise those values.

If a belief is held entirely in isolation, without *any* implications for the way the person holding it will act, then I suggest that those people holding it do not

---

<sup>17</sup> Ibid., 176

<sup>18</sup> Ibid., 176

<sup>19</sup> Ibid., 167 [emphasis added]

form a group in any helpful or interesting sense, and Gilbert cannot intend to include them. There are no social or political implications, for example, if a number of people believe that oak leaves are blue when no-one is looking at them, provided that that belief does not in any way change their pattern of life, even to the extent of trying to persuade others to agree (which would count as a purpose).

Having thus laid out Gilbert's narrow intentionalist view, the next stage is to criticise her account. It is, I suggest, lacking on three levels. Firstly, it is an error to treat social groups and collectivities as one and the same. Secondly, Gilbert's criteria, although sufficient to identify one kind of collectivity with which we, as political philosophers, should be concerned, are not necessary. Thirdly, her condition is unnecessary for the constitution of even a *social group*. The relationship between collectivities and social groups is more complex than that. As far as the wider project of this thesis is concerned, the relevant point is the second. I will, therefore, discuss the other two, which I have made at more length elsewhere, only insofar as is necessary to make my case with regard to that one.<sup>20</sup>

### III. Three objections to Gilbert

This section will criticise the Narrow Intentionalist Thesis as unnecessarily restrictive. It is immediately vulnerable to two objections (call them the Involuntary Members and the Unintentional Members Objections), both of which identify intuitively compelling cases that Gilbert's model fails to accommodate. Her later work allows her at least some scope to evade the former, but in attempting to avoid the latter, she falls foul of a third, the Circularity Objection.

The Involuntary Members Objection accuses Gilbert of implausibly excluding groups which are standardly considered (some of the most important) collectivities, through her implicit incorporation of the idea of voluntariness.

The Narrow Intentionalist Thesis' *joint readiness* has distinct overtones of voluntariness. Gilbert (1989) spells out: 'One might say that someone who expresses quasi-readiness to do A in C in effect *volunteers his will* for a pool of wills to be set up so that in certain circumstances, that pool will be dedicated to a

---

<sup>20</sup> Cripps 2005



certain end.<sup>21</sup> Moreover, on her account, a set of individuals constitute a plural subject if and only if each of them can appropriately use the English word “we”, in a ‘full-blooded’<sup>22</sup> sense, which also requires some kind of expression of willingness.

*[The] appropriate full-blooded use of “we” in “Shall we do A?” requires that each of the people referred to has in effect expressed to the others his willingness to share with the others in doing A, in circumstances of the type at issue.*<sup>23</sup>

How, then, is such an account to accommodate cases where members are born into a group, notably the nation state but also various tribes, or even families? There is no sense in which all the individuals considered to be members of a nation state have chosen willingly to join it.

Gilbert (2005) clarifies herself out of this particular problem, by specifying that expressions of readiness need not be entirely voluntary. She counts coerced agreements as genuine agreements (although, as with coerced promises, they are not necessarily agreements which, all things considered, one is morally obliged to uphold).<sup>24</sup> By the same token, even coerced expressions of joint readiness (for example, expressing readiness for the joint commitments of the political society into which one was born, when the only alternative is deportation<sup>25</sup>) are enough for membership of a plural subject, although without necessarily rendering the individual morally obliged to fulfil the obligations thereby incurred.<sup>26</sup>

She defends this viewpoint by pointing out that, once there is an established practice of making agreements, ‘agreements may be made in situations in which one party is, indeed, pressuring the other to enter the agreement’. The point is that the agreeer is still *willing* to agree in what Gilbert calls ‘the *intention* sense’.<sup>27</sup> Unlike someone who signs a paper without realising what it says and so agrees without intending to do so, the person who signs with a gun against her head is knowingly agreeing, although she is not acting voluntarily.<sup>28</sup> This seems to fit

---

<sup>21</sup> Gilbert 1989, 198

<sup>22</sup> Ibid., 200

<sup>23</sup> Ibid., 179

<sup>24</sup> Gilbert 2006, 75-82 and 228-9

<sup>25</sup> Gilbert’s example, which she discusses in the context of actual contract theory. (2006, 76)

<sup>26</sup> Gilbert 2006, 235

<sup>27</sup> Ibid., 76

<sup>28</sup> Gilbert’s example (2006, 76)

our standard understanding of agreements. Explaining what happened later, the person is likely not to say: “I didn’t agree – he put a gun to my head”, but rather: “I agreed, but I had no choice – there was a gun to my head.”

In such circumstances, Gilbert argues, the agreeer is in a sense obligated: she has, by the act of agreeing, committed herself to some joint endeavour. However, in another sense (‘a sense standardly connoted by the qualifier “moral”’<sup>29</sup>), she is not, because she was coerced.

This line of argument, if successful, could save Gilbert from the Voluntariness Objection. Nonetheless, there remains the related question of *how* exactly members of (say) nation states or families are supposed to have expressed their readiness to be jointly committed with the other members. Those who are born into a state rarely have to make some declaration, on coming of age, to participate in the joint endeavours of their country. Nor, in countries where voting is not mandatory, can they be said to be declaring their willingness to participate in the political process by doing so in this explicit way.

Gilbert allows for expressions of joint commitment which are not explicit, and which evolve over time. Thus, she would presumably count remaining in the country, paying taxes, obeying the laws and so on as an expression of readiness, made to others in society, to do one’s part. This still leaves some potentially problematic cases, however. Suppose someone is born into a country but, on coming of age, refuses to obey its laws or pay its taxes. He ends by being taken forcibly to prison, where he never gives up trying to escape. At no point, coerced or otherwise, has he made any expression of readiness for joint commitment. Does this exclude him from membership of the nation state? It is not intuitively obvious that it does. (Later sections will make it clear how my model can accommodate such examples.)

Perhaps Gilbert would interpret his being in the country at all, benefiting from an infrastructure (at the very least, roads and water systems) built with public money, as an expression of readiness. This, however, seems to stretch the notion of expression of readiness almost to emptiness. Moreover, there are individuals, whom we intuitively count as members of some central collectivities, who not only do not, but cannot express joint readiness in any intentional sense at all.

---

<sup>29</sup> Gilbert 2006, 81

This line of argument can be taken further still: to challenge not only the need for “expressed joint readiness” but also for “awareness”. This is the Unintentional Members Objection. Against the Narrow Intentionalist Thesis, we can point to individuals intuitively regarded as members of some central collectivities, who cannot be (or are not) aware of the common purpose or goal. Against the Broad Intentionalist Thesis, which requires that all members regard themselves as part of a group, we might point out that such individuals do not consider themselves as being so.

Consider the following special case individuals, familiar from the second set of examples identified in Section I. The cases are formulated in terms of the family, but the arguments could also apply at a wider level, to tribes or nations.

The first is the baby. Babies, intuitively, are members of families. Families, intuitively (and on my account), are collectivities. Gilbert wants them to count as such.<sup>30</sup> However, babies are not aware of being part of a collectivity. They are not aware of the common purposes of a family (that is, all the various things a family does together). Nor would we want to deny them their status: the rest of the family, after all, considers them to be members and they are tied to the other members through the interests which they have in common.

One possibility is to amend the Narrow Intentionalist Thesis to require that individuals, if they are not currently aware of the goals and their being shared, will at least become so. However, this (itself a significant amendment of the thesis) would not accommodate the next problematic case: the severely mentally handicapped person. The exclusion of the mentally handicapped child from a family is as repulsive a result as the exclusion of the infant. The natural response would be to say that all individuals within the set who can be aware of the goals/purposes (aware that they are shared, etc) will be so aware, but that there may be some who are not so able. This, which would involve going further still from the core idea of the intentionalist account, would nonetheless leave untouched the third problematic case: the teenage rebel.<sup>31</sup>

The rebel is not a baby, nor is he mentally handicapped. He would deny being part of the family and resent any demands made on him by the family. Nonetheless, it seems on most intuitive readings that he is still part of the family,

---

<sup>30</sup> Gilbert 1989, 234

<sup>31</sup> Thanks to David-Hillel Ruben for the example.

and his parents still think he is. To include such cases, the Narrow Intentionalist Thesis would have to be amended yet again, to allow for members of a collectivity who *could* be aware of the common goal, purpose *or interest* (aware that it is a common goal, purpose or interest) but who deny this.

The above discussion has taken us further and further from the Narrow Intentionalist Thesis. Full membership of families (or tribes or nations) by any of these three individuals is incompatible with the view that '[t]he core type of group membership is at some level intentional – it is not acquired unwittingly, such as by inheritance simply; social groups [collectivities] involve a substantial kind of unity [and] core group members will perceive that such unity exists'.<sup>32</sup>

Gilbert (1989) insists that her account can be extended to include (at least some of these) 'complex cases'. While maintaining that plural subjecthood is 'central to collectivity existence as intuitively conceived', she does not want to exclude the baby from the tribe or family. She also wishes to allow individuals to become members of clubs (as, indeed, they do) 'knowing little about [the club's] aims and ideology'.<sup>33</sup> Thus, she introduces the notion of entry rules.

Plural subjects proper can develop along a variety of lines. For instance, a plural subject may adopt a name; it may develop a sense of specialness and a desire for exclusivity: 'We don't want anyone else in on this, unless we all really like her'. Or it may have a sense of need for new members. For such reasons the existing members may establish rules that determine who may become a member. The possibility of such entry rules allows for complex cases. The rules can in principle be so capacious as to allow those lacking the ability to be members of the initial plural subject to be considered members none the less. Thus infants can be thought of as 'members of the tribe', though they have no conception of the tribe as a whole. Similarly, an adult may know nothing of the tribe but still be a member according to the group's rules.<sup>34</sup>

Unfortunately, Gilbert leaves us unclear what limits there are on these entry rules. As it stands, this paragraph seems to say that anyone can become a part of any collectivity, simply because those individuals who make up the plural subject at its core so decide. Suppose two of my young cousins decide to form an

---

<sup>32</sup> Gilbert 2006, 168

<sup>33</sup> *Ibid.*, 233

<sup>34</sup> *Ibid.*, 233

Orlando Bloom fan club. They then feel a 'sense of need for new members'<sup>35</sup> and so set up entry rules such that all of their female first cousins count as members. My sister and I become members of a collectivity without knowing it, or wanting to be. (They could equally well set up the rules so that everyone in the world was a member.)

This cannot be what Gilbert intends. Rather, she wants to allow cases such as the person who joins a club without really knowing much about what it does – that is, who espouses a goal without specific knowledge of what exactly that goal is. This point could be accommodated by suggesting that entry rules apply to those who *want* to be part of the collectivity. However, this fails to let in exactly that other set of cases with which we were originally concerned. The baby and the mentally handicapped person cannot (we assume) *want* to join the collectivity, because they have no idea what it is all about. The teenage rebel does not want to consider himself a member. We need some way of identifying those individuals who can be part of the group: those who, in each case, the supposed plural subject core is *entitled* to consider as members.

Later, Gilbert appeals to the idea of imputed membership, as a kind of stop gap until full, intentional membership can be assumed.

[Core members] may stipulate that others, for instance, their children, are to be regarded as members for some or all practical purposes. These others then have imputed membership. They will not be core members or members proper unless they come to participate in the joint commitment that constitutes the political society in question.<sup>36</sup>

This could be plausible with regard to children, who can be expected to achieve core membership of their family groups and societies in time. However, it implies rather repugnantly that mentally handicapped individuals can never have more than imputed membership of their families or states, and would similarly exclude cases such as that of the teenage rebel.

Moreover, Gilbert's attempts to amend her account leave her vulnerable to the Circularity Objection.<sup>37</sup> According to her paradigm case, for a set of individuals to constitute a collectivity, they must all be in the appropriate "we" mental state.

---

<sup>35</sup> Ibid., 233

<sup>36</sup> Gilbert 2006, 239

<sup>37</sup> Sheehy also accuses her of circularity (2002, 386-9)

This allows her to define social groups in terms of the thoughts and beliefs of member individuals. However, by introducing notions such as an entry rule or imputed membership, she is adding a social concept to the conditions on the existence of a collectivity (or, according to her own account, a *social* group).

Gilbert says disappointingly little about what entry rules would be, but they would presumably permit an individual to become part of a social group, without awareness of shared purposes, because some kind of criteria have been agreed to by those already in the group, according to which they count as members. This, certainly, is the idea behind imputed membership. Thus the concept of an entry rule is not, as she appears to suppose, innocent of anything social, but requires its own analysis in terms of collective ideas, and so cannot be thrown in at this stage.<sup>38</sup>

#### **IV. Accommodating tough cases**

Having thus attacked the Gilbertian model, where do we stand? The Broad Intentionalist Thesis has been rejected. At least some members of those groups which we standardly regard as central collectivities do not see themselves as such. Moreover, having chipped away at the Narrow Intentionalist Thesis, we are left with the idea of a number of individuals who have a shared or common goal or purpose (or, as with the teenager example, an interest in common), but who have not each necessarily entered readily into this shared end, and not all of whom can (or do) even acknowledge it.

In other words, we have come to something very close to the definition given in Section I:

A set of individuals constitutes a collectivity if and only if those individuals are mutually dependent through some common or shared purpose, goal, or ATC interest, whether or not they acknowledge it themselves.

---

<sup>38</sup> As Section V will highlight, such “counting” of something as significant in a sense over and above its material properties is central to the realm of the *social*.

As we have seen, this accommodates all three of the problematic inclusions above. The baby is mutually dependent with his parents (and siblings) for the achievement of the goals of family life (broadly, living peacefully and happily together). Of course, the baby himself does not have *goals* – he is unable to formulate them. However, the dependence is “mutual” because, although the conscious, practical effort in pursuit of these goals will be made by the others, it is not only part of *their* shared aims that the baby be included, and be happy, healthy, not screaming all the time, etc., but this healthy family atmosphere is in *his* interest. A similar argument can be offered for the mentally handicapped family member.

We can also assume, plausibly, that there are various ways in which the teenage rebel is tied to the family. Perhaps he still inhabits the same house as his parents and siblings, perhaps his parents want to spend time with him and they all, including him, benefit from a stable family atmosphere. It seems that there are common *interests* in play: fundamental interests in a stable, healthy family life, which can only be met together. (I will come back to this question of interests.)<sup>39</sup>

The non-intentionalist model can also accommodate a nation such as Spain, its collectivityhood hinging not on *acknowledgement* of shared purposes but simply on mutual dependence. This is a key aspect of the defence of this model. However, it requires further explanation. It is not entirely straightforward to identify the common interest, goal or purpose through which *everyone* in that state, including rebel individuals parallel to the teenager, is mutually dependent.

A first suggestion would be that the common goals or purposes of the Spanish are the many things they want to achieve as a nation, from winning the World Cup to defending themselves (if necessary) against attack. However, various individuals within Spain will not (for example) support Spain in the football. Perhaps their parents were Argentinean and they support Argentina instead.

This is not in itself a difficulty, so long as there are *some* common goals, purposes or interests at the national level: there are intuitively many different collectivities within the nation, including cultural subgroups. However, some

---

<sup>39</sup> Note, however, that there are circumstances in which the teenager could conceivably cease to be part of the collectivity, although he would still be part of the biological family: for example if he deliberately, physically separated himself from his family, moving city and setting up an entirely independent existence

individuals will not even regard themselves as committed to the most general goals, such as that of defending Spain against enemy attack. Consider extremist Basque separatists who do not regard themselves as part of Spain at all, and actively work against the national government.

Even in these cases, I argue, there is a common interest in the relevant sense. Those who live and work under a government are mutually dependent through the myriad of institutions in the country. Their long-term interests in keeping safe, healthy and educated (of generally getting through life in a worthwhile way) can, under the circumstances, only be pursued together. To use an analogy, consider the Tragedy of the Commons, familiar from game theory.<sup>40</sup> There may not be any fellow feeling between the owners of the animals, but they are in a situation such that it is in all of their long-term interest not to destroy the land, and they can only achieve this by co-operating and agreeing to limit the total number grazing on it.

Even if the terrorist subgroup denies this, it is in everyone's (including the terrorists') ATC interest to find a peaceful solution. (The underlying claim – that genuine opportunities for security and health are of such fundamental importance to a flourishing human life that it cannot be in an individual's ATC interest to put them at serious risk – will be defended in the next chapter.) Of course, it is possible that the solution would ultimately be a separation into two nations, and two collectivities, one of which was no longer a subgroup of the other. However, that does not change the current mutual dependence of the individuals.

This has defended the non-intentionalist model against two of the objections levelled more successfully against Gilbert's intentionalist view. To show that the model can also avoid the third, it is necessary to go into more detail on the relationship (and the distinction) between the social and the collective.

## **V. Defending the distinction 1: social properties, groups and aggregates**

There is no space here, nor is it necessary, to elucidate all the many rival accounts of what exactly is at stake when we speak of social properties.<sup>41</sup> It is

---

<sup>40</sup> See Chapter Four, Section IV for further discussion.

<sup>41</sup> I discussed social properties in more detail in my MPhil thesis (Cripps 2005, 36-41). See Ruben for one carefully thought-out example (1985, 110-7)



enough to draw on illustrative examples to offer a draft of a necessary criterion so general as to be uncontroversial.

The following are clear cut examples of social properties being true of something, or someone. Gordon Brown is the British Prime Minister. Italy is a member of the European Union. A certain piece of plastic and paper in my bag is my UCL student card. As these illustrate, an uncontroversial necessary criterion for a property's being *social* is that given in Section I: the property has a significance over and above the material or physical properties with which it coincides, and has this because of the recognition by a certain group of individuals that it will have it. (I am trying to capture the idea of a general understanding that doing such-and-such a thing will *count as* doing x.)

Using this intuitive idea, we can bring out the threefold distinction, highlighted in Section I, between: collectivities (sets of individuals of which at least one collective property in the strong sense is true), social groups (roughly, collectivities whose collectivityhood hinges on the applicability of some social property, but I will return to this in Section VI), and social aggregates (sets of individuals who share a social property but do not constitute a collectivity). On this account, it is neither necessary nor sufficient for collectivityhood that some social property be true of some set of individuals.

This gives rise to an obvious objection: that I have unnecessarily complicated matters, both in distinguishing between collectivities and (their subset) social groups, and in defining social groups other than through shared social properties. In response, beginning with the latter, I will attempt to rebut the suggestion that it is sufficient to establish social grouphood (and so collectivityhood<sup>42</sup>) that a set of individuals share some social property.

Why should sharing a social property be any different to sharing some physical property (for example, having blue eyes), unless having that property ties the individuals in some further way? (As with, for example, being black, or being a woman: those who share such properties acquire shared interests, including an interest in not being discriminated against.)

Take, for example, all those people who happen to be entitled to a Spanish passport (a social property). Many of these (in particular those living in Spain) do

---

<sup>42</sup> It would be an intuitively repugnant account of collectivities which actually *excluded* some social groups.

constitute an intuitively important collectivity. But it is probable that there are some individuals, scattered across the world, who are entitled to a Spanish passport (for example by virtue of a parent who emigrated as a child) but consider themselves, and indeed are, in no way specially attached to Spain. They are unconcerned about its performance in the World Cup, and upset by terrorist attacks on Madrid no more than they would be by such attacks anywhere else.

This wider set of individuals is not tied in a way which is either politically relevant or which should excite the interest of social scientists looking to explain the results of human interaction. There is no reason why such a set should constitute a collectivity, any more than those with blue eyes.

The non-intentionalist account can accommodate this intuitively important point. As the previous section indicated, many of those entitled to a Spanish passport (because they are part of the Spanish nation state) are tied through shared goals and interests. However, those individuals whose parentage entitles them to a Spanish passport but genuinely have no special interest in Spain, are not so tied. Thus, the wider set constitutes only a social aggregate.

## **VI. Defending the distinction 2: collective v. social properties**

A shared social property is insufficient for collectivityhood. As we have seen, Gilbert offers a more structured account of a collectivity, taken (by her) as the same thing as a social group, which incorporates the element of *acknowledged significance* central to the social. However, as we have also seen, she is unable to account for certain key members of social groups without falling foul of the Circularity Objection. Instead, my model appeals to the idea of the *collective*, independent of any preconceived views about the social. This point can be made by starting, as she does, with the idea of individuals doing something together, but understanding it more broadly.

There seem to be three ways in which a number of individuals might bring about some result between them. The first is *pure coincidence* – a combination of individual acts which, by chance, take place at the same time and so have some particular effect. An example would be if five individuals, going about their separate lives, happened to step on the five trigger points required for a bomb to

go off, at exactly the same time. Such cases, being one-off and unpredictable, can be ignored for the purposes of this project.

The second is *predictable combined effect*: the combined outcome of individuals acting independently in pursuit of individual goals. Suppose a primary school is surrounded by houses which are being refitted. All the owners have decided independently to embark on the projects. The noise of construction from one house would not be a problem, but the combination is so loud as to render teaching extremely difficult. Such cases, I contend, are politically interesting under certain circumstances: where the combined outcome is harmful, in a morally significant way, to some other individual or individuals. I will explain this idea at the end of this chapter.

The third possibility is *collective achievement of a common end*. It is around such cases that the politically interesting understanding of a collectivity revolves. It is here that we have a prudential case for special obligations between members. Moreover, Gilbert concentrates on this third case. However, careful attention reveals two sub-cases, corresponding to two understandings of “common end”, the second of which Gilbert (with her focus on the *social*) overlooks.

The first case involves a genuinely shared goal: it is the goal of the individuals to do something *together*, as a set. (Recall the example from Section I of the friends climbing Monte Rosa.) In the second case, action has to be taken by a number of individuals, as a group, in order for any individual’s individual goal to be achieved, *even if each individual is indifferent as to whether the others achieve their goals*. The *Lost* case is an example; another is the classic Prisoner’s Dilemma, where it is in everyone’s interest to find a workable means of co-operating to reach the second best outcome for each, rather than leave each individual to make a selfish decision, which results in a worse outcome for all.<sup>43</sup>

In *both* cases, it is possible to identify a set of individuals who are mutually dependent for the satisfaction of their purposes and goals. This is the reasoning behind the definition offered in Section I: a collective property in the strong sense is true of them (that is, they constitute a collectivity). (Note that I use “in the strong sense” to distinguish these from predictable combined effect cases, where a weaker collective property, corresponding to the weak collective harm

---

<sup>43</sup> See Chapter Four, Section IV for further discussion of the Prisoner’s Dilemma

discussed in Chapter One – would be applicable. I will come back to this in Section XI.) This is clearly distinct from the idea of a social property sketched above. However, it leaves us able to account for social groups as a subset of collectivities, as outlined in Section I.

A collective property is true of a set of individuals whenever they are mutually dependent in the relevant way. Their being so may be a result of entirely material factors (as with the Commuter Village or Escaping Prisoners) or it may be a result of social factors (as with the friends on Monte Rosa: there is an additional *recognised significance* to their reaching the top, together).

Social groups are those sets of individuals who constitute a collectivity on the back of social properties. This could happen in two ways: it could be because of some social property or social properties true of the individuals that they constitute a set of which a collective property is true (as with the Spanish, as will be discussed below), or it could be an integral part of the common goal that at least one social property becomes or remains true of someone, some set of individuals, or something. (Consider, for example, the supporters of a second division football team, whose shared goal includes various social properties – winning the FA cup, getting promoted and so on – becoming true of that team.)

This leaves us with an account on which the attribution of a social property is neither necessary nor sufficient for collectivityhood. Section V showed why it is not sufficient. The defence of it as unnecessary is twofold. Firstly, we have seen from Gilbert that, in defining collectivities *as* social groups (i.e. including this element of *recognised significance* in the necessary criteria for collectivityhood), one runs the risk of circularity. This account, as I will spell out below, avoids this problem whilst still accommodating such central collectivities as nation states.

However, in avoiding this problem we have reached a model of politically (and social scientifically) interesting collectivities which is much broader than many would intuitively accept. Sections VIII, IX and X will defend this result.

## **VII. Avoiding circularity**

My non-intentionalist account, in separating the collective from the social, is able to accommodate the Spanish nation state as both a collectivity and a social group without falling foul of the Circularity Objection.

Gilbert tries to define the *social* in terms of the thoughts and beliefs of individuals (“we” thoughts). However, as we saw in Section III, she introduces the notion of an entry rule, which is a social property, and thus runs the risk of circularity. For me, there is no such problem, because my more objective definition is of a *collectivity*, which is not a social concept. This distinction allows me to claim that a collective property may hold true because of social factors, but that does not make it any less a collectivity. Rather, it also makes it a social group.<sup>44</sup>

How is this done? As Section IV showed, the inhabitants of the Spanish nation state count as a collectivity because they are bound together (at the very least) by their fundamental interest in a peaceful, secure existence. Because the mutual dependence arises as a result of historical social arrangements, the Spanish are not only a collectivity but a social group. But, regardless of the status of “being Spanish” as a social property, a person born in Spain finds herself, *as a matter of fact*, in relations of mutual dependence through common interests with all those with whom she actually shares her nation state. That is, she is a member of that collectivity. Thus, the account avoids circularity.

## VIII. Controversial collectivities

In order to avoid excluding certain individuals from collectivities, I have defined collectivityhood according to actual rather than universally acknowledged mutual dependence. Avoiding circularity, it has been argued that collectivityhood does not rest on the attribution of *social* properties. I have also appealed to the basic idea of doing or achieving something together, to identify cases in which individuals might be mutually dependent with regard to goals, purposes or interests, in a way independent of the definition of the “social”. This section will address the more controversial implications of the non-intentionalist model, by returning to the non-intentional collectivities identified in Section I. I will then further defend the claims of even such controversial collectivities to the attention of political philosophers and social scientists.

---

<sup>44</sup> Using Graham’s example of the Unintentional Clique. I have argued elsewhere that it is possible to have not only a collectivity but even a social group which is characterised by unintentionality on the part of *all* of its members. However, for the purposes of my project here this stronger claim is unnecessary. (Cripps 2005, 74, and Graham 2002, 72-3)

Two of the controversial examples follow naturally from the second of the two senses of collective achievement of a common goal, laid out in Section VI. That is, where action has to be taken by a number of individuals, as a group, in order for any individual's individual goal to be achieved, even if each individual is indifferent as to whether the others achieve their goals.

Firstly, the house-owners in the Commuter Village do not consider themselves as part of a group, although they can only keep the green tidy (which each wants, on purely selfish grounds) if all contribute to this end. This is a borderline case. While it is involuntary, and non-intentional in the broad sense, it is at least arguable that there is intentionality to the extent that the individuals, if they use their intelligence, will know that they are mutually dependent.

Secondly, recall the Escaping Prisoners, rowing a boat none would be able to move alone. In Sheehy's original example, which does not specify that the prisoners are delusional, it does seem that each individual, while he is not voluntarily espousing the situation and does not think of himself as a member of a group, must at least be aware of the situation. This would render the collectivity intentional in the same sense as in the Commuter Village. However, in my adapted example, each of the prisoners believes himself to be possessed of superhuman strength, such that he could propel the boat across the water himself. On my objective account, the prisoners still constitute an *ad hoc* unintentional collectivity. As a matter of *fact*, they are mutually dependent: escape, for each, can only be achieved through the action of all.

In the Rival Gangs Case, each individual would vehemently refuse to co-operate and deny any kind of collectivityhood with the others. However, the discussion of the baby, the mentally handicapped family member and the teenage rebel has already demonstrated the need to expand the criteria for collectivityhood to include mutual dependence through certain key individual *interests* even if they are not acknowledged as goals by the individuals themselves. (The next chapter will fill this out by defending a capabilities-based model of human flourishing.) The individuals in this case are mutually dependent in exactly this way: co-operation is the only way to preserve each one's fundamental interest in staying alive.

This leaves us with the most controversial example: Searle's pack of hyenas hunting a lion.<sup>45</sup> Here, there is mutual dependence without the capacity to acknowledge a genuinely shared goal. I would argue that the hyenas constitute a collectivity. However, the key arguments of this thesis will not in any way rest on the inclusion of such cases, so there is no need to defend this claim in detail. It was mentioned to raise the (to me, interesting and advantageous) possibility of expanding my model to accommodate non-human animal cases.

### **IX. Meeting Graham's conditions**

This section will complete the defence of the non-intentionalist model, by appeal to work by Graham. Graham and Sheehy are both non-intentionalists. However, unlike me, they do not question Gilbert's assumption that social groups and collectivities are essentially the same thing. Moreover, rather than offer a rival model of the circumstances under which a set of individuals constitute a collectivity (as with my appeal to mutual dependence), they attack the intentionalists by appeal to certain explanatory or descriptive criteria which a collectivity should meet, and which are satisfied by some non-intentional collectivities.

Sheehy claims that groups are individuated by their ineliminability from 'our best explanations in the social domain'.<sup>46</sup> He asserts this "ineliminability" for groups including his escaping prisoners. I have argued elsewhere that, understood as a claim about irreducible causal impact, or explanatory autonomy, this is flawed: social groups have an irreducible causal impact, in that causal explanations which cite the social properties of social groups are not reducible to explanations in terms of only the individual properties of individuals. However, those collectivities which are not social groups, do not have this kind of autonomy.<sup>47</sup> I omit these arguments here, as it is not necessary, for the purposes of this thesis, to engage in detail with the explanatory or causal status of the *social*.

---

<sup>45</sup> Graham 2002, 71-2, citing Searle 1995

<sup>46</sup> Sheehy 2001, 117

<sup>47</sup> Cripps 2005, 36-58

Graham does not require explanatory autonomy. Rather, he argues, a number of individual human beings form a collectivity (or social group) if and only if they satisfy certain criteria:

- i) *that they act in ways whose significance can be adequately captured only by an ineliminable reference to some corporate body as part of which they are acting, where*
- ii) *what that corporate body does is distinct from anything which they as individuals do, and*
- iii) *where the corporate body is a persisting one whose survival is relatively indifferent to the persistence of the particular individuals which compose it at any particular moment.*<sup>48</sup>

I will argue that all the collectivities on my non-intentionalist account, including those which are not social groups, will satisfy these conditions. This requires a certain reading of the conditions, which may not be entirely in line with Graham's own intention. However, the point to be made is: other than by *starting* with the assumption that only social groups should be allowed to qualify for collectivityhood, there is no reason to interpret the conditions in such a way as to exclude other collectivities.

(Note, moreover, that I would eschew the word "corporate", with its implications of corporate action – a rather stronger notion than collective action – and would replace it with "collective". This is an important move beyond Graham and one which I will defend in my treatment of condition (ii).)

Because he has rejected intentionalism, Graham would deny that the "significance" to which he refers in condition (i) must be in terms of how the individual thinks of himself as acting (that is, as part of a group). However, this leaves us puzzled as to what exactly is involved. What Graham means seems to be something like the significance of the individual in his example of a mob.<sup>49</sup> Clearly, the combination of all the actions of all the individuals in those particular circumstances has a certain effect, and it is this effect to which we refer in accounting for the "significance" of the individual actions. However, left as that, any individual action which, however coincidentally, contributed to an

---

<sup>48</sup> Graham 2002, 68-9

<sup>49</sup> Ibid., 72.



overall effect, could be said to have “significance”. (Recall the three levels of doing something together, in Section VI.)

At the other extreme, Graham might require the kind of significance incorporated in the rough draft of a necessary condition for a *social* property – being understood by some group of people to “count as” something else.<sup>50</sup> However, this seems unlikely: Graham himself acknowledges that there is an additional significance to individual acts, even in coordinated action cases, which they could not have outside that context.<sup>51</sup> This essentially social reading is not the only possible understanding of “significance”, and to understand it in this way would simply be to assume that all collectivities are also social groups. Moreover, given that he regards collectivities and social groups as one and the same, he could be accused of circularity.

I suggest that the significance lies in the fact that the individual acts (intentionally or otherwise) contribute to a wider effect which results from certain relations in which the individuals stand. That is, relations such that the individuals are mutually dependent for the achievement of certain of their purposes, interests or goals: those ends, “common” in either of the two senses identified in Section I, can only be achieved, for all the individuals, by the set as a whole.

It still remains to specify what is meant by “ineliminable”, which will be done by engaging with arguments by Graham Macdonald and Philip Pettit. In line with Graham himself, my contention is that the reference is “ineliminable” in the sense brought out by the notion of *expressive* or *descriptive* (rather than *explanatory*) autonomy. To simplify a definition given by Macdonald and Pettit, being able to refer to expressively autonomous collectivities ‘enables us to give expression to truths that we cannot formulate [without such terms]’.<sup>52</sup>

---

<sup>50</sup> Sections I and VI

<sup>51</sup> ‘[S]uppose I am a member of the town’s garrison and leave it in circumstances where others do so too, so that the town’s defence collapses in a way which would not be possible if I alone had left. Or I act with a number of others to lift some heavy object which I could not possibly lift on my own... In these cases, I act as an individual in coordination with a number of other individuals and the significance of what I do is thereby transformed.’ (Graham 2002, 67)

<sup>52</sup> Macdonald & Pettit 1981, 115. More precisely, ‘one sort of entity X exists over and beyond another sort Y if and only if... the addition of terms by means of which we refer to X-type things enables us to give expression to truths that we cannot formulate in a language with terms for referring to Y-type items’.

Drawing on Macdonald and Pettit's case for the expressive autonomy of institutions, it is possible to argue that even the most controversial collectivities on my account are expressively autonomous. Macdonald and Pettit argue that, for many of the sentences formulable in a language with concepts for referring to institutions and individuals,

there seem to be no accessible counterparts in the truncated language which is deprived of the means of referring to institutions... The sentences from the richer language contain expressions with which no expressions from the more austere one can be more appropriately aligned.<sup>53</sup>

Consider the difficulty of finding individualistic counterparts to group predicates such as "has members" or "regroups".

However, there is a possible objection, from what Macdonald and Pettit dub the 'expressive individualist'.<sup>54</sup> That is, 'that since groups are composed of people, anything that involves a group involves the people who are its members, and... any predicate therefore which attaches to the group must do so only insofar as a corresponding predicate attaches to its members.'<sup>55</sup>

Macdonald and Pettit respond by drawing a distinction between two possible claims glossed together in the objection, one of which is accurate, the other of which does not follow. The first idea is that the composition of groups is such that any group predicate is supervenient on individualistic predicates (by which they mean, put most roughly, that group properties cannot change without some change in individual properties). The second is that there will be a specific individualist counterpart available for any given group term. This is not guaranteed by supervenience, which is, rather, consistent with a group predicate applying 'in circumstances characterised by any of an open-ended list of individualistic predicates, or predicates compounds, so that we are not in a position ever to identify a disjunction of such terms which applies just when the group predicate does so'.<sup>56</sup> (This is a version of the multiple realisations argument.)

---

<sup>53</sup> Macdonald & Pettit 1981, 118

<sup>54</sup> Ibid., 119

<sup>55</sup> Ibid., 119

<sup>56</sup> See Macdonald & Pettit 1981, 119.

We use group terms without considering which particular individualistic predicates apply in each case, nor would it necessarily help us to do so. As Macdonald and Pettit point out, '[t]here seem to be an indefinite number of individualistically describable ways in which group predicates may be realised'.<sup>57</sup> For example, our ability to say of a group that it "gains members" does not depend on whether we can link the terms to individualistic characterisations of people, even if we can show which characterisations, in a particular case, were behind the application of the group predicate.

Given, then, that appeal to a group predicate 'enables us to detect a kind of phenomenon that escapes our conceptual grasp if we are restricted to the narrower range of concepts',<sup>58</sup> Macdonald and Pettit grant expressive autonomy to institutions. Similarly, all the collectivities identified by the model defended in this chapter have expressive autonomy because there will always be multiple combinations of individual action which could produce the collective result. Consider, for example, the numerous different combinations of individual acts by crew members which bring about the sailing of a ship.

To make the two points (ineliminability and significance) together, recall the four controversial collectivities. In the Commuter Village, many combinations of individual action could lead to the green being kept tidy and pleasant. Some individuals might pick up litter while others simply refrain from spreading it, or vice versa; one person might turn his music down when he has a barbecue, another change out of his spiked boots to play football on the green, or vice versa. Moreover, while it is true that each individual refrains from making noise, etc, because he wants to keep the green pleasant *for himself*, it is nonetheless the case that, because of the relations in which the individuals find themselves (of mutual dependence through a goal which can only be pursued together) their acts only lead to the satisfying of these goals (keeping the green clean) in combination. This brings out the significance of their behaviour.

The same can be said of the hyenas. Indefinite numbers of different combinations of individual actions can lead to the group result. One particular hyena might grab the lion and three others form a line so as to stop it from escaping, or some other way around, or some completely different combination

---

<sup>57</sup> Macdonald & Pettit 1981, 120

<sup>58</sup> *Ibid.*, 120

of actions. In all cases, the behaviour of each individual hyena does not achieve the goal (death of the lion) except as one of a number of coordinating acts. In this sense, its significance is only captured by ineliminable reference to the collectivity.

The other two cases are more problematic, but still meet the condition. With Sheehy's prisoners, the significance of each individual's action (rowing) cannot be brought out without considering the fact that that action, *in combination with* the actions of the other escaping prisoners, will enable them to escape. But is reference to the group ineliminable in the relevant sense? At first sight, it might seem that because each individual is doing one thing (rowing) that contributes to the overall result, the multiple realisations argument does not apply. But this is not the case. Even given that all are rowing, there are numerous different combinations of precise individual actions which could yield the overall result. For example, all might row with force  $r$ , or two might row with force  $r-s$  and the other two with force  $r+s$ , and so on.

This leaves the Rival Gangs Case. Again, there is a significance to an individual's action – a significance *for him* – which can only be brought out by consideration of context. In fighting the others, he is contributing to a situation in which his own interests are undermined, because, in this small-scale state of nature, co-operation is required for survival. Moreover, the outcome (whether disaster or co-operation) could come about in innumerable ways, whether  $x$  bashes  $y$  over the head, and  $z$  retaliates, or  $y$  attacks  $x$  and  $z$  runs away and hides, or some other combination, the outcome at the collective level is the same: failure to co-operate resulting in disaster.

Now consider Graham's other two conditions. A slightly amended (ii) (what that *collective* body does is distinct from anything which they as individuals do) will hold for the model of collectivities defended in this chapter. However, this does reinforce the distinction between my understanding of the conditions and Graham's. As he notes, the difference (between what the collectivity does and what the individual does) can be contingent rather than necessary. However, his examples all involve *social* properties. He points out that in some cases (for example, when it comes to electing a president) it is impossible for an individual to do what a collectivity does, while in others it is only in a certain context that

the individual cannot do what the collectivity does. For example, only a jury can find a person guilty in certain criminal trials, but there are some civil trials in which the judge alone can do so.

With collective properties, on my account, it is enough that the individual cannot *physically* do what the collectivity does. For example, a crew working collectively sails a large ship across the water; the orchestra working collectively plays a symphony.

Graham appears to exclude such combinations by drawing a distinction between such cases of *coordinated* human action, and what he calls collective action, which involves a corporate entity.<sup>59</sup> However, his attitude is puzzling. As already noted, given that he himself acknowledges the *significance* of coordinated action,<sup>60</sup> it seems he must regard his elaboration of the further conditions – (iii) and in particular (ii) – as at least contributing to his defence of the distinction between coordination and collective action. But, again, presumably because he has started from the idea of social groups and collectivities as synonyms, he seems simply to *assume* that the only relevant cases to consider are those involving social properties. This rather begs the question.

Such a limitation is, I contend, unnecessarily restrictive given Graham's own ambitious starting point. His project, broadly put, is to explore the notion of how we *should* behave, that is, to discuss normative rather than explanatory reasons, and how this is influenced by the interlinkedness of our lives. As he puts it, he is concerned with 'how the reasons we have for acting are affected by the simple thought that we are *social* creatures. Whenever we act, either singly or in collaboration, we act together.'<sup>61</sup> However, when he formalises his starting point into 'two truths', there is no mention of the "social".

These "truths" are: firstly, that 'everything a person does carries causal implications for the lives of other human beings', and, secondly, that 'some of the things a person does gain their significance from being part of some collective action, carried out by a number of human beings'.<sup>62</sup> By taking the social, almost before he starts, to delineate the interesting or relevant realm

---

<sup>59</sup> Graham 2002, 67

<sup>60</sup> Ibid., 67. See also Footnote 51

<sup>61</sup> Graham 2002, 1

<sup>62</sup> Ibid., 1

within which we act collectively, he is excluding some of the cases covered by those “truths”, and so restricting the realm (and so the moral or political implications) of his enquiry.

Recall, again, the controversial examples. In each, there is a clear (if sometimes contingent) distinction between what the collectivity does and what the individual does. The hyenas, not any individual hyena, kill the lion. The prisoners, as a collectivity, escape: none could do so alone. The gang members, between them, destroy their chances of survival. The Commuter Village, as a collectivity, keeps the green tidy.

Finally, consider a slightly amended condition (iii): survival of the *collective* body is relatively indifferent to the persistence of the particular individuals which compose it at any particular moment. This requires some clarification. There are two possibilities as to what Graham means. He could mean that the collectivity must be indifferent to the persistence of all the individuals in it at any one time. But, while some collectivities appear to meet this condition (boat race crew, university), there are also some obvious counterexamples (a group of friends, a couple). Of course, even with the former cases where, over time, all the members can change and the group remain, it is far from clear that it would be the same group persisting over time if all the members left at once and were replaced by new members. However, these are Theseus’ ship type problems, as acknowledged by Graham himself, common to metaphysics, and not unique to collectivities.<sup>63</sup>

The other option is that he requires at least some members to be able to leave and be replaced. In this case, the condition is more widely satisfied. Recalling some of the examples above, one home owner in the Commuter Village could sell his house to a newcomer, without the collectivity ceasing to exist, there could be one gang member more or less on the island without the mutual dependence which renders them a collectivity ceasing to bind those who were there, and individual hyenas could drop out of, or join, the hunt.

However, there still seems to be a problem. Graham notes that ‘[p]erhaps the smallest [collectivity might be constituted] by just a couple’.<sup>64</sup> But it looks as

---

<sup>63</sup> Graham 2002, 70. For a discussion of such cases in metaphysics, see, for example, Wiggins, 161-71

<sup>64</sup> Ibid., 76

though the removal of one person from a couple does effectively put an end to that collectivity. Similarly, would the collectivity of a group of three friends continue if one of them left? In the first case, we can immediately rebut the objection by pointing out that one person alone cannot be a collectivity. Moreover, in general it seems at least possible that few-person groups could sustain some changeover. A family is still a family when an extra child is born, or one parent dies. In the friends case, it is harder to sustain this argument. However, it is at least plausible that a small friendship group would persist if one member left and was replaced (compare a three-member pop band such as the Sugarbabes with its turnover of members).

We can accommodate this by allowing for borderline cases. Note, importantly, that these are borderline only in the same way as we can cite borderline cases with regard to artefacts. How many legs does one remove from a chair before it ceases to be a chair?

Thus clarified, condition (iii) is satisfied. Moreover, it brings out an important distinction between *ad hoc* and lasting groups. Some collectivities on my account – those such as Sheehy’s prisoners that I have already dubbed *ad hoc* – will not satisfy (iii). However, lasting collectivities will. It is these that are of particular interest to practical philosophers, policy makers and social scientists.

## **X. Two objections to the non-intentionalist model**

The previous section defended even the more controversial collectivities accommodated by the non-intentionalist model, as satisfying the criteria for collectivityhood laid out by Graham. This completed the defence of the model, and its expansion of the understanding of a collectivity beyond that of a social group. However, there are two obvious, if conflicting, objections to be tackled: that the conditions for collectivityhood laid out here are too weak, and that they are too strong.

Let us begin with the latter. This contends that it is unnecessary to require mutual dependence of *all* individuals on *all* individuals within a collectivity. Take the example of the nation state. Surely, the objector points out, everyone in the state can get along very well without *everyone* paying taxes, so long as enough people do so to support the health, security and education institutions?

The response to this is twofold. Firstly, I have already acknowledged the possibility of collectivities of which only a subset is active. For example, women and the suffragettes. In such cases, there is some goal which, if achieved at all, would be achieved for all. This could be because it is a genuinely shared goal (each woman genuinely espouses the goal that all women should have the vote), or simply because of the nature of the good (votes for women means votes for you if you are a woman, and it is in each woman's interest to have a vote). The mutual dependence clause is met because each individual's achieving her goal (or interest) depends on all the others also having theirs achieved or satisfied, but much or all of the positive *effort* may be put in by only the active subset.<sup>65</sup>

However, secondly, the situation is different when we are considering goals or interests which do not by their nature, but in the circumstances do, need to be achieved by all for all. In this case, membership of the collectivity is contingent both on individuals having the interest (or goal or purpose) and on co-operation being required for its achievement. The nation state example seems to fall into this category.

On further examination of the example, it becomes possible to reject the objection, which narrows down unnecessarily the ways in which an individual could co-operate. *Omissions*, as well as actions, can count. Thus, while not everyone has to pay taxes for a peaceful, secure environment to be provided for all, this common interest does require everyone to refrain from committing terrorist attacks, running around yelling abuse at people on the street, or putting poison in the drinking water, and so on.<sup>66</sup>

This leaves us with the claim that the non-intentionalist model is too weak. The objection claims that the model counts as a collectivity *any* set of individuals

---

<sup>65</sup> I will come back to this distinction as part of Chapter Five's derivation of individual duties.

<sup>66</sup> Clearly, this implies that not all our so-called political duties can be grounded (directly) prudentially in this way. This is unsurprising. There are various points to note. Firstly, prudentially grounded special responsibilities to one another (i.e. prudentially grounded duties to co-operate in achieving the end which bind a collectivity together) do not come automatically with a mandate for coercive imposition. Such a mandate could be added by moral considerations, either in order to prevent harm to others, as with morally salient potential collectivities (see next section) or through a variant of the principle of fairness which legitimates coercing individuals into co-operating with a collective entity if that is required to satisfy their own fundamental interests, or if the costs of their noncooperation to other members were too high. (See Chapter Five, Sections III, IV and VIII) Legitimate coercion might further require a condition not only about how decisions were made (democratically, say, but with basic rights protected for all) but also about how duties were allocated, for example requiring that all rather than just some contribute through the tax system. However, this is beyond the scope of this project.



who would benefit, in terms of some interest or goal they have, from co-operating with one another. But this means that, because (say) you and I would both benefit from being fitter, and going running together each morning would help us to do so, we count as a collectivity. This is counterintuitive.

This objection misrepresents the model. There is a distinction between potential mutual advantage and mutual dependence. According to the understanding of mutual dependence incorporated in the non-intentionalist account, a set of individuals constitute a collectivity only in the following three situations.

Firstly, most uncontroversially, a set of individuals constitute a collectivity if they have an acknowledged shared goal in the simplest sense of wanting to achieve particular end together: as with the friends whose aim it is to reach the top of Monte Rosa (and get down again safely) *as a group*.

Secondly, they constitute a collectivity even without this strong sense of an acknowledged shared goal, if co-operation is *required* for the achievement of each individual's acknowledged goals or purposes. Thus, suppose several individuals have got themselves, independently, half way up Monte Rosa. They are faced with bad weather and each wants to get down. However, they need to rope together and use all the equipment they have between them to do so. Then, they constitute a collectivity even if none cares about getting down *together*. This can be distinguished from the case in which each could get down perfectly well alone but it might be pleasanter (or faster) to combine forces. Then, there would be potential for mutual advantage but, unless all choose to adopt the new shared goal (getting down together), they would not constitute a collectivity.

Finally, even where there is no such mutual dependence for the achievement of acknowledged individual goals, a set of individuals can constitute a collectivity if co-operation is required for the satisfaction of each individual's ATC interest. This is not the same as hinging collectivityhood on each individual having *an* interest in something that would require co-operation for its satisfaction. Thus, to use Robert Nozick's example, the fact that I could be said to have an interest in a public address system for my street, which would require

co-operation with other residents for its achievement, is not enough to render them and me a collectivity.<sup>67</sup>

Rather, the idea is that there are some things (for example, life, health, bodily integrity) which are of fundamental importance in any human life. Nozick can plausibly argue that, while he might have *an* interest in the public address system, it is not in his ATC interest because there are other goals he considers more important, or other, more fundamental, interests which outweigh it. However, it would not be in his ATC interest to be permanently deprived of one of these more central components of a flourishing life, whatever he thinks. This claim will be filled out in the next chapter, when a capabilities model of human flourishing will be defended. For now, the point is simply that mutual dependence for the satisfaction of an interest which is not also espoused as a goal or purpose is not enough to ground collectivityhood, unless that interest is an individual's ATC interest in not being deprived of a central capability.

Returning to the example, suppose the climbers do not espouse the goal of getting down safely: some want to go on to the summit despite the changed conditions. They might still constitute a collectivity but only if: a) their lives are at serious risk if they don't descend immediately, *and* b) the situation is as described above, so that none has the equipment or expertise to get down alone, but all could do so together. If the risk of completing the ascent were only slightly increased, or if some could descend independently, they would not be mutually dependent in the required way.<sup>68</sup>

This should clarify that the account is not too weak. Note also that it avoids a problem faced by Gilbert. Section III asked how Gilbert could place limits on her entry rules (or imputed membership) – the mechanism to which she was obliged to appeal to include babies or the mentally handicapped in her collectivities. The risk would be that anyone could become a part of any collectivity, simply because those individuals who make up the plural subject at its core decreed this to be so. (Recall my example of the Orlando Bloom fan club set up by two of my young cousins.)

---

<sup>67</sup> Nozick 1974, 93-5. See also Chapter Five, Sections III, IV and VIII

<sup>68</sup> Although those thus capable might nonetheless have a positively grounded moral duty to co-operate in getting everyone down. (See Chapter Five, Section IV)

This, as we have just seen, is not a problem for an account which draws the boundary according to actual mutual dependence through goals, purposes or ATC interests. To recall another example, parents and sibling can legitimately consider their baby (or adopted baby) part of the collectivity which also includes them. However, they cannot arbitrarily set entry rules such that the three of them plus (say) Cruz Beckham constitute a collectivity. Equally, my cousins cannot arbitrarily include my sister and me in their collectivity, by introducing entry rules. We do not share their goal, nor is actively admiring Orlando Bloom necessary to our flourishing as human beings.

The two girls are not entitled to claim that they constitute a collectivity even with all other Orlando fans in the world. Each fan has the aim of adoring him in whatever ways she (or he) chooses, but that does not make them mutually dependent through a common goal. It is neither the case that all fans (as with these two) have as part of their goal that it be carried out with the others, nor that it is a goal which can only be carried out *en masse*. This is in contrast to the teenage rebel, whose family are entitled to consider him part of the collectivity even if he denies it. This is because he *is* tied to them, simply as a matter of fact, by his (and their) fundamental interests.

## **XI. Morally salient potential collectivities**

Having completed the defence of the non-intentionalist model laid out in Section I, I have one addition to make to that model. I promised to return to the second sense of alleged collective action identified in Section VI. That is, the predictable combined effect of individuals acting independently in pursuit of individual goals. This suggests a way in which individuals could constitute at least a morally relevant set, without having any *common* goals or interests, in either of the senses used in my definition.

I offered the example of the householders round the primary school. Now, I suggest that the set of individual homeowners in this situation has true of it a collective property *in the weak sense* and is, therefore, a *potential collectivity*. The individuals are in a situation such that, were they to adopt or acquire certain goals or interests, they would become a fully fledged collectivity. Were the

owners to take on the goal of eliminating the disruption, they would have to do so *collectively*.

Why should political philosophers care about such cases? Because there might be collective responsibility for harm, in the sense identified in Chapter One: harm resulting from the predictable aggregation of individual actions. Potential collectivities who are collectively responsible (in the weak but morally significant sense) for harm to some person or persons are, I suggest, *morally salient potential collectivities*. In such cases, I have argued that, assuming moral impermissibility, there is a collective duty to redress the harm. Put in the terms of this chapter, the potential collectivity should turn itself into an actual collectivity with the goal of ending (or at least mitigating) the collective harm. There is thus often a case for treating these particular potential collectivities, at a political level, as though they were actual collectivities. (I will return to this point in Chapter Five.)

## **Conclusion**

I have argued that the common ground between political philosophy and philosophy of social science requires reworking, so as to start, in the former, with a broader model of a collectivity than that offered by the intentionalists. On this non-intentionalist model, a set of individuals constitute a collectivity if and only if those individuals are mutually dependent through some common or shared purpose, goal, or ATC interest, whether or not they acknowledge it themselves.

A set of individuals also constitutes a social group if its collectivityhood hinges on the attribution of *social* properties, that is, if it is because of some shared social property that the individuals are mutually dependent or if it is an integral part of the common goal, ATC interest or purpose of the collectivity that social properties become or remain true of some person, persons, group or object. Social properties are properties which are *taken* (through the acknowledgement of a number of people) to have a certain significance over and above the material and physical properties with which they coincide.

I have thus argued against Gilbert's model of a collectivity, taken to correspond to the Narrow Intentionalist Thesis, and taken the counterexamples to her account as the starting point for the development of a more objective, non-

intentionalist model, on which the set of social groups is a subset of that of collectivities.

Before closing, it is helpful to bring out the difference between our two projects. Gilbert's primary focus, at least in her later work, is the *obligations* which we acquire as members of certain collectivities (in particular, political obligations acquired by members of states to one another). Given an exclusive focus on that side of our relationship with fellow citizens, it is at least *prima facie* understandable to consider only (or at least primarily) those with whom we are intentionally bound: that is, those who will fulfil similar obligations.

However, there are two sides to the membership of states, and the other one concerns those *to whom* we acquire responsibilities. Many of these will also participate in the political process. But what about those (children, the mentally handicapped, arguably – although this goes beyond this particular project – sentient non-human animals) whose interests are as bound up with ours as each others' are, but who cannot actively participate in any collective decision making, or knowingly bear the burdens of membership of a state?

The intentionalists could, in excluding such individuals from full membership, intend them nonetheless to receive the protection of the state. But such "half way" or "second class" membership, especially for mentally handicapped adults, is unsatisfactory, and open to dangerous interpretation in practice.

In fact, when we consider the actual situation, the dependence (as with the infant and the family) does to a great extent go both ways. Children, the mentally handicapped and even sentient animals are in a sense *subject* to the laws of the land, even if they do not knowingly comply with them: their behaviour is restricted in accordance with them. My model eliminates the contingent, add-on, element of their membership and allows for the possibility that, even if they cannot *participate* in collective decision making, they are at least *represented* in it: that their individuals interests are taken into account, perhaps through some specific individual or group who speaks on their behalf.

### **Chapter Three**

#### **In the Human Interest: Defending Capabilities**

What is required for a human life to go well?

I am doubly required to answer this dauntingly fundamental question. Chapter One, defending an expanded understanding of collective responsibility for harm, tacitly assumed the availability of an account of morally significant harm. To identify the key components of a flourishing human life (a human life which can be said to be *going well*) is also to identify certain aspects, interference with which constitute harm to an individual.

Chapter Two defined a collectivity as constituted by a set of individuals mutually dependent through some common or shared goal, purpose, or *all-things-considered (ATC) interest*. I contend that political philosophy should be concerned with collectivities in this broad, non-intentionalist sense, rather than restrict its discussion of collectively-acquired duties to acknowledged social groups. The specification of mutual dependence not only through goals or purposes but also through ATC interests calls for an account of what can be said to be in a person's interest and how I can justify appeal to certain interests even if the individual does not acknowledge them as her goals or purposes.

More generally, this chapter can be understood as providing a basic moral grounding for my project, by responding to the question: what, if anything, has moral value? Such value might be assigned to the flourishing of communities and societies, over and above that of the individuals who make them up; to the well-being of non-human animals, to certain religious ideals, to the natural world, or even to certain objects of aesthetic value, such as great works of art. These are more or less controversial.

However, I am taking as foundational only the uncontroversial claim that the flourishing of an individual human life has moral value. That is not to say that I deny the moral value of all of the other ends suggested above. Indeed, I believe that at least the flourishing of non-human animals has value, and the arguments of this thesis could ultimately be extended to accommodate this. But by starting with only the more limited claim, I am starting from a truth which is (more or less) universally acknowledged, and so can only be strengthening my argument.

That said, in order to make any progress from this unproblematic claim, it is necessary to answer a much more controversial question. That is, precisely the question which I am already required to address: what exactly is meant by “the flourishing of an individual human life”? This chapter draws on the idea of *capabilities to function*, in particular as elucidated by Nussbaum, to provide an answer.

However, there is a crucial objection to the capabilities approach, as identifying the elements of (or prerequisites for) a flourishing human life: that it must fall, unavoidably, into one of two camps: undiluted preference satisfaction, or unacceptable paternalism. Call this the Preferences-or-Paternalism Puzzle (PPP) and the two components of the complaint the Preferences Accusation and the Paternalism Accusation, respectively.

It is essential that I respond to this accusation. On the one hand, I have defended an extension of the notion of a collectivity to include sets of individuals bound together by ATC interests which are not necessarily their acknowledged goals. If the capabilities approach turns out to collapse into undiluted preference satisfaction, then I cannot use it, as an understanding of human flourishing, to defend this distinction. On the other hand, given that I will later be drawing *political* conclusions, in terms of individual and collective duties, from such collectivityhood, it is essential to demonstrate that, in avoiding undiluted preference satisfaction, the model is not unacceptably paternalistic. There will thus be two stages to the defence against a charge of paternalism. One is to defend the capabilities account, *as a model of human flourishing*, which will be done here. The other is to defend my use of the account as providing a notion of our individual interest which can be used to ground certain collective and individual duties. The arguments of this chapter go some way towards this, but will be supplemented in Chapter Five.

Another significant objection has been made to the capabilities approach. Ronald Dworkin has argued that, taken as a metric for distributional equality, it collapses into either equality of welfare, or his own (or Rawls’) equality of resources.<sup>1</sup> I will also address this, for two reasons. Firstly, because of its potential implications at the wider level (that is, for the capabilities approach as

---

<sup>1</sup> Dworkin 2000, 299-303

identifying prerequisites for human flourishing). Secondly, because Dworkin's argument, if successful, could threaten any attempt to apply the capabilities model *politically*, and I need to show that it does not threaten mine. While I do not here defend *equality* of capabilities or even a capabilities-based account of distributive justice, I do rely on a version of the approach in grounding some special responsibilities, and I at least refuse to rule out the possibility that such responsibilities, in certain cases, could legitimately be coercively enforced.

The argument will be structured as follows. Section I will give some background to the capabilities approach, and lay out the version to be defended. This is something like Nussbaum's list of central human functional capabilities.<sup>2</sup> Sections II and III will argue that Dworkin's attack turns out either to be unconvincing, or to rest on the PPP. I will then offer a solution to the PPP, beginning in Section IV by presenting an understanding of the capabilities approach from which it follows that it is *not* in someone's interest, because it undermines the possibility of a fully flourishing human life, to be permanently deprived of a capability. This draws a dividing line between the capabilities approach and a subjectivist welfare account.

Section V will defend the approach against the Paternalism Accusation by drawing a distinction between acceptable and unacceptable paternalism. I will argue that a "vague thick"<sup>3</sup> account along the lines of Nussbaum's list is not *unacceptably* paternalistic. Finally, Section VI will tackle one outstanding objection.

## **I. Human flourishing: the capabilities approach**

The capabilities approach, inspired by Aristotle, pioneered by Amartya Sen and spelled out by Nussbaum, has huge intuitive appeal. Its central claim, as Sen puts it, is:

that for many evaluative purposes, the appropriate "space" is neither that of utilities (as claimed by welfarists), or that of primary goods (as demanded by Rawls), but that of the *substantive* freedoms – the capabilities – to choose a life one has reason to value.<sup>4</sup>

---

<sup>2</sup> Nussbaum 2000b, 78-80

<sup>3</sup> Nussbaum 1990, 217-28

<sup>4</sup> Sen 1999, 74 [emphasis added]



The approach is Aristotelian in that it draws on the notion of human flourishing as more substantive than preference satisfaction. Modern moral philosophy has already drawn the idea of flourishing to provide a non-expressivist ground for judgements of *moral character*. The evaluation of human behaviour, as Philippa Foot puts it, is part of practical rationality, but abstract rationality alone is not enough. Rather, ‘the evaluation of human action depends also on essential features of specifically human life’.<sup>5</sup>

Foot ties moral virtues (and defects) to ‘what human beings are and what they do’ by appeal to the same kind of “Aristotelian necessity” as is invoked in identifying as a “good” lioness one who provides food for her cubs and teaches them to hunt.<sup>6</sup> As part of this account of goodness in human behaviour, she defends a deep notion of happiness, tied to the idea of living a “good” or flourishing life. ‘To flourish,’ she says, ‘is here to instantiate the life form of that species’.<sup>7</sup> We find deep happiness in ‘things that are basic in human life, such as home, and family, and work, and friendship’.<sup>8</sup>

The capabilities approach is similarly concerned with what it means to live a flourishing life, but, unlike Foot, does not attempt to marry the ideas of a virtuous human and one who is living a “good” or flourishing life. Rather, the approach has standardly been concerned with evaluating human lives for the purposes of guiding political action. Within the “equality of what?” debate, appeal to capabilities is intended to steer a middle way between an equality of welfare which mandates financing expensive tastes, and a Rawlsian or Dworkinian equality of resources, which is insufficiently sensitive to different abilities to translate resources into lives worth living.<sup>9</sup> Recent work has developed the approach as a way of identifying for policy purposes which people

---

<sup>5</sup> Foot 2001, 14

<sup>6</sup> Ibid., 15. See also Foot 1978b, 132-147

<sup>7</sup> Foot 2001, 91

<sup>8</sup> Ibid., 88. She adds: ‘These are, in a way, ordinary things, even though, in a person such as Wittgenstein, the chief joy of his life was in the quest for truth, and in other exceptional men and women it is in artistic creation or the exploration of strange lands.’ (Ibid., 88) This acknowledgement of different conceptions of the good or rational plans of life accommodates, as Nussbaum’s list does, the centrality of autonomy, or meaningful choice, in a valuable human life.

<sup>9</sup> See Rawls 1971, Sen 1980 and 1992, Dworkin 2000, 11-119. Note that this section of *Sovereign Virtue* (Chapters One and Two) is reprinted with only minor corrections from Dworkin. “What is Equality? Part One: Equality of Welfare”, *Philosophy & Public Affairs* 10 (1981): 185-246 and “What is Equality? Part Two: Equality of Resources”, *Philosophy & Public Affairs* 10 (1981): 283-385

are “disadvantaged”,<sup>10</sup> and Sen has defended it as a tool for development economists and other decision-makers at the international level.<sup>11</sup>

For the purposes of this project, what matters is that the approach offers a way of identifying what, broadly, is in a person’s interest, over and above what they espouse as a goal. Individuals could, accordingly, be said to be tied to one another in politically relevant ways which they did not intend and might not acknowledge. The approach also lends substance to the notion of doing morally significant harm to some person or persons (the kind of harm, as I will discuss in Chapter Five, which policy-makers might legitimately interfere to prevent).

In all these cases, however, the question arises: on what is this evaluative tool based? The issue, according to Sen, ‘is... the capability to live really long (without being cut off in one’s prime) and to have a good life while alive (rather than a life of misery and unfreedom) – *things that would be strongly valued and desired by nearly all of us*.’ A paragraph later, he stresses the importance of ‘having more freedom to lead *the kind of lives we have reason to value*’.<sup>12</sup> So are these things (these “freedoms” or capabilities) important *because* nearly all of us desire them, or is it our having *reason* to value them that does the normative work, in which case the question becomes one of how to fill out what we have reason to value?

Nussbaum goes further than Sen in that she lays out a provisional list of central human functional capabilities. I will adopt a model of the central capabilities based broadly on this list. Similar questions, arise, however, as to whether is she is taking a *moral* rather than a purely *political* view. I will take the approach as making a substantive claim about what is valuable in a flourishing human life, and one which can be defended in a certain way. This, I suggest, is compatible with at least one reading of Nussbaum, although my arguments need not rest upon its being so.

Nussbaum is ambiguous. Her article for *Liberalism and the Good* appears to make a substantive moral claim. She bases the list of capabilities on a sketch of ‘*The Constitutive Circumstances of the Human Being (or: the Shape of the*

---

<sup>10</sup> See for example, Wolff and De-Shalit 2007

<sup>11</sup> See especially Sen 1999

<sup>12</sup> Sen 1999, 14 [emphasis added]

*Human Form of Life*)<sup>13</sup> and specifies allegiance to the Aristotelian social democracy camp. However, *Women and Human Development* assigns the capabilities a status similar to that of the Rawlsian political conception: as ‘the object of an *overlapping consensus* among people who otherwise have very different comprehensive conceptions of the good’.<sup>14</sup>

Her remarks elsewhere provide clarification. In “The Good As Discipline, the Good as Freedom”, she affirms the list as ‘something that people from many different traditions, with many different fuller conceptions of the good, can agree on as the necessary basis for pursuing their good life’<sup>15</sup> but also distinguishes between Rawls’ “thin” theory and her own more comprehensive theory of the good.<sup>16</sup> In “Aristotle, Politics, and Human Capabilities” she acknowledges ‘a significant shift’ towards political liberalism, and in her understanding of the political status of the capabilities. She remains wedded to a ‘substantive-good approach’<sup>17</sup>, but clarifies that hers is a basic-level ethical *judgement* on which there can be deep and broad consensus – a “provisional fixed point”<sup>18</sup> in a process of reflective equilibrium. She maintains her link with Aristotle by arguing that he was not ‘deriving ethical norms from metaphysical biology’<sup>19</sup> either.

Thus, it seems plausible to take her capabilities, unlike Rawls’ primary goods, as specifying what *is* involved in a flourishing human life (not just a political proxy for different individual views). However, her source for the information – her evidence as to what those capabilities actually are – is neither metaphysical nor biological, but rather broad consensus at a deep level among actual human beings (backed by empirical research) despite differences in the *detail* of different conceptions of the good.

This gives substance to Sen’s claim that the capabilities (or freedoms) are what we have reason to value (we have reason to value them because they *are* the key ingredients of a flourishing human life) and renders it compatible with

---

<sup>13</sup> Nussbaum 1990, 219

<sup>14</sup> Nussbaum 2000b, 5. See Rawls 1993 for an extensive discussion of overlapping consensus and its role in political liberalism.

<sup>15</sup> Nussbaum 1998, 318

<sup>16</sup> I will elaborate on this discussion in Section V.

<sup>17</sup> Nussbaum 2000a, 117

<sup>18</sup> Ibid., 120

<sup>19</sup> Ibid., 118

his describing them as what most of us would want (they are, which is how we are able to identify them as the central elements of a flourishing life). Moreover, while on the face of it such a dissociation from “biology” appears to move away from Foot’s species-based understanding of flourishing, Nussbaum’s appeal to consensus at a deep level is not so very far from Foot’s reasoning in the following elucidation of her notion of “deep” happiness.

We are tempted to think of deep happiness as explicable psychologically *in a way that makes it possible to separate it from its objects*. But why should this be possible? Why shouldn’t the communality of meaning not depend here on a shared reaction among human beings to certain things that are very general in human life. Are not these reactions shared even by people of very different cultures; not, of course, exactly, but nevertheless with sufficient similarity for people of one age or culture to understand depth of happiness over a birth and depth of grief about the death of parent, child, or friend?<sup>20</sup>

Having thus outlined the reasoning behind it, we turn to Nussbaum’s list of central human functional capabilities.

1. Life  
Being able to live to end of normal human life, not die prematurely/before life reduced so not worth living.
2. Bodily health  
Being able to have good health (including reproductive health), have adequate shelter/nourishment.
3. Bodily integrity  
Being able to move freely from place to place, have bodily boundaries treated as sovereign, have opportunities for sexual satisfaction/choice over reproduction.
4. Senses, imagination and thought.  
Being able to use senses, imagine, think and reason in “truly human” way, cultivated by adequate education, to experience and produce self-expressive works and events, to use mind in ways respected by freedom of expression (political and artistic), to search for meaning of life in one’s own way, to have pleasurable experiences/avoid non-necessary pain.
5. Emotions  
Being able to have attachments to things/people outside ourselves, to love and grieve, feel longing, gratitude and justified anger, not to have emotional development blighted by fear/anxiety/abuse or neglect.

---

<sup>20</sup> Foot 2001, 88

6. Practical reason  
Being able to form conception of good and engage in critical reflection about planning of one's own life.
7. Affiliation
  - (A) Being able to live with/toward others, to recognise and show concern, interact socially, to imagine the situation of others and to feel compassion; capability for justice and friendship.
  - (B) Having the social bases of self-respect and non-humiliation, being able to be treated as dignified being of equal worth to others, being able to work as human being, exercising practical reason/enter into meaningful relationships of mutual recognition with other workers.
8. Other species  
Being able to live with concern for and in relation to animals/plants/world of nature.
9. Play  
Being able to laugh/play/enjoy recreational activities.
10. Control over one's environment
  - (A) Political – being able to participate effectively in political choices governing life, having right of political participation, free speech and association.
  - (B) Material – being able to hold property (land and movable goods, in terms of real opportunity, having property rights on equal basis with others, right to seek employment on equal basis with others, freedom from unwarranted search and seizure.<sup>21</sup>

Except insofar as I will refer back to it in defending this kind of account against the charge of paternalism, it is not necessary to examine or defend this list in detail. Rather, it is enough to say: it is a prerequisite to flourishing that individuals have access to certain vaguely defined central functional capabilities. Indeed, the key interests on which I will rely, in later chapters, are less controversial than some included on Nussbaum's list.

Having laid out the capabilities account, we can now turn to the first stage of the argument.

---

<sup>21</sup> Nussbaum 2000b, 78-80. [Text abbreviated] Wolff & De-Shalit propose the following additions: 11. Doing good to others (Being able to care for others as part of expressing humanity. Being able to show gratitude); 12. Living in a law-abiding fashion (Being able to live within the law, not being forced to break the law, cheat or deceive other people or institutions); 13. Understanding the law (Having a general comprehension of the law, its demands and the opportunities it offers to individuals) (2007, 50-1. Text abbreviated.)

## II. The Dworkinian complaint 1: resources & preference satisfaction

Dworkin asks: ‘Is Sen’s own positive conception of equality – equality of capabilities – really different from equality of resources? If it is, is it really different from equality of welfare?’<sup>22</sup> He answers in the negative. On one reading, he argues, Sen’s account collapses into equality of welfare; on the only other reading (which he thinks Sen intends), it collapses into equality of resources.

There are three preliminary points to make. Firstly, I neither pretend nor need, for the purposes of this thesis, to provide a complete answer to the question of whether *equality* of capabilities is an appropriate political goal. Even holding capabilities (rather than resources or welfare) to be the appropriate metric for distributive justice, it would be possible to defend, as does Nussbaum, a sufficiency or threshold-based model of distributive justice.<sup>23</sup>

Secondly, Dworkin’s complaint is concerned with redistribution but the capabilities approach, as I take it, is much broader than that. It is concerned with what is needed for a flourishing human life. There are, therefore, two issues to consider in examining Dworkin’s argument. In terms of the wider debate, to what does his criticism of the capabilities approach amount? And does it succeed in the narrower context of the “equality of what” discussion? My concern is primarily with the former. However, these sections will also have implications for the latter.

Thirdly, Dworkin’s case is made against Sen, who does not offer a list of capabilities. However, he would presumably offer a similar criticism of any attempt to use Nussbaum’s list as a metric according to which politicians should seek to make people “equal”.

Sen’s viewpoint, as Dworkin quotes him, is:

A person’s achievement... can be seen as the vector of his or her “functionings”, consisting of beings and doings, [varying] from such elementary things as being adequately nourished, being in good health, avoiding inescapable morbidity and premature mortality, etc., to more complex achievements such as being happy, having self-respect, taking part in the life of the community, and so on. The claim is that functionings are

---

<sup>22</sup> Dworkin 2000, 301

<sup>23</sup> For example, Nussbaum 2000b, 75

constitutive of a person's being, and an evaluation of well-being has to take the form of an assessment of these constituent elements. Closely related to the notion of functionings is that of the capability to function. It represents the various combinations of functionings (beings and doings) that the person can achieve.<sup>24</sup>

There are, Dworkin thinks, two possible readings of this paragraph. One is 'as suggesting that people should be made as nearly equal as possible in their capacity to realise the "complex" achievements of happiness, self-respect, and a significant role in the life of the community'<sup>25</sup>. The second is that government 'should strive to insure that any differences in the degree to which people are not equally capable of realising happiness and the other "complex" achievements should be attributable to differences in their choices and personality and the choices and personality of other people, not to differences in the personal and impersonal resources they command'.<sup>26</sup>

I will begin with this second reading, which Dworkin takes to be closer to Sen's intention,<sup>27</sup> and on which, he argues, equality of capabilities collapses into equality of resources. I will contend that his claim, taken at the level of the wider debate, amounts to a version of the Preferences Accusation.

On Dworkin's model, "equality of resources" is achieved in two steps: external resources (wealth and income) are equalised; then deficiencies in internal resources (strength, health, talent, intellect, and so on) are compensated for by adjusting the allocation of external resources according to what insurance would, hypothetically, be taken out against them.<sup>28</sup> The motivating idea of "equality of capabilities", by contrast, is to render people equally capable of achieving certain key functionings.

There has already been considerable academic debate around the question of the distinction (or lack thereof) between the two models. This has tended to focus on interpersonal variation in the ability to convert resources into capabilities to function, either because of natural differences, or because of social or material

---

<sup>24</sup> Sen 1992, 39-40, quoted by Dworkin 2000, 301

<sup>25</sup> Dworkin 2000, 301

<sup>26</sup> Ibid., 303

<sup>27</sup> Ibid., 302

<sup>28</sup> Ibid., 65-119

factors.<sup>29</sup> I will not go over this ground, except to note that Dworkin's response would be to argue that his notion of internal resources allows him to accommodate internal differences between individuals, such as handicaps, but that this would still leave him open to the point that his "equality" of resources is compatible with dramatic differences in actual capabilities.

Consider Andrew Williams' example of capabilities discrepancies resulting from other people's choices. Ann and Bob, according to Williams, do not have equal capability 'to combine parenthood and occupational success' because of the difference in the general willingness of men or women to take up the counterpart role in the family.<sup>30</sup> Dworkin acknowledges similar counterintuitive implications in his hypothetical case of shipwreck survivors allocating resources on a desert island, but responds that 'the contingent facts of raw material and the distribution of tastes are not grounds on which someone might challenge a distribution as unequal. They are rather background facts that determine what equality of resources, in these circumstances, is.'<sup>31</sup> To treat them otherwise, he claims, would be to fall back into equality of welfare, which he sees himself as having effectively refuted.

However, such a response is convincing (insofar as it is) only so long as the two models, equality of welfare and equality of resources, are taken to be exclusive alternatives. It is undermined if a plausible metric for equality can be offered which can accommodate cases such as Williams', without collapsing into equality of welfare. That is, precisely what equality of capabilities would claim to do. This will be revisited in the next section.

The argument to be offered here takes a different line. It hinges on the role of individual choice in determining with respect to *which* capabilities individuals should be rendered equal. Dworkin cites the reading of equality of capabilities which he sees as collapsing into his model as calling for no objective ranking of "functionings". Rather, he says:

It is enough to distribute impersonal resources equally, and find devices, like the hypothetical insurance market, to mitigate differences in personal resources so far as this

---

<sup>29</sup> See Sen 1992, Dworkin 2000, as well as, for example: Williams 2002; Pogge 2004; Pierik & Robeyns 2007.

<sup>30</sup> Williams 2002, 31

<sup>31</sup> Dworkin 2000, 69-70



is possible. Then we can allow people, through their choices against that more nearly equal background, to make their own rankings of “functionings” important to them.<sup>32</sup>

However, in talking of “functionings” rather than “capabilities”, Dworkin neglects a key distinction within the capabilities approach. His idea is that his own resourcist model can satisfy the demands of the capabilities approach, in this reading, by providing individuals with resources such that they can pick and choose across the *functionings* important in living a full human life. But the capabilities approach treats not functionings, but *capabilities*, as the appropriate metric for distributive justice. Thus the key question in establishing a convergence must be whether the two accounts can agree on whether it is sufficient to provide individuals with resources such that they can “make their own rankings” of capabilities.

Recall that a central element in Dworkin’s resourcist model is the use of hypothetical insurance to calculate compensation (in external resource terms) for internal resource deficiencies. He explains:

If... everyone had at the appropriate age the same risk of developing physical or mental handicaps in the future... but the total number of handicaps remained what it is, *how much insurance coverage against these handicaps would the average member of the community purchase?* We might... say that but for (uninsurable) brute luck that has altered these equal odds, the average person would have purchased insurance at that level, and compensate those who do develop handicaps accordingly.<sup>33</sup>

However, basing calculations on the insurance the *average* person would take is only a practical compromise. Dworkin specifies in a footnote that, ideally, equality of resources would compensate for personal resource deficiencies according to the level of insurance the relevant *individual* would take out. Averaging, he says, ‘is a simplifying assumption only, made to provide a result in the absence of the detailed (and perhaps... indeterminate) information that would enable us to decide how much each handicapped person would have purchased in the hypothetical market. If we had such full information, so that we

---

<sup>32</sup> Ibid., 303

<sup>33</sup> Ibid., 78 [emphasis added]

could tailor compensation to what a particular individual in fact would have bought. the accuracy of the program would be improved.<sup>34</sup>

Suppose, then, that we were in a world of such refined technology and information that distributions could be calculated, in line with this ideal version of equality of resources, by appeal to the insurance that each individual, hypothetically, would have taken out against having the internal resource deficiency which he or she happens to have. Suppose, further, that an individual lacks some internal resource the absence of which prevents her from exercising a central capability. However, this individual does not have a preference for the relevant capability, and so would not have taken out insurance against lacking it.

The footnote from Dworkin quoted above makes it clear that this ideal world equality of resources would not provide her with that capability. Indeed, to do so would appear to conflict with the core idea of equality of resources, structured as it is around the idea of an envy test. A person would not be compensated using hypothetical insurance for the lack of a capability which she does not herself value, because she would not envy anyone else's bundle of resources simply because they were able to exercise that capability and she was not.

Is this enough for equality of capabilities? Only if the capabilities approach, too, would target equality only in those capabilities that the individuals themselves valued. (Or, to put it another way, if it would seek only to provide an overall allocation of capabilities which met the envy test.) Dworkin says that the idea, common to the two approaches, is to 'allow people, through their choices against that more nearly equal background, to make their own rankings of "functionings" important to them'.<sup>35</sup> But to support his conclusion that equality

---

<sup>34</sup> Ibid., 478 (Footnote 5 to Chapter 2 of *Sovereign Virtue*) Note, however, that even if this "ideal" equality of resources did appeal to *average* insurance, Dworkin's claim of collapse would not hold. He would have a better chance of achieving a merger between the two approaches, but only because the capabilities approach could provide a rationale – for appealing to average insurance – which would otherwise be lacking. With allocations determined in this way, most individuals would regard themselves as over or underinsured. Dworkin's model would thus involve the, as it stands, arbitrary allocation of a resource bundle, objective in the sense that it does not correspond to the individual's own valuations. There is no rationale, in ideal resourceist (envy test) terms, for doing this. However, by basing a resourceist system on a capabilities account of flourishing, it could be rationalised. Dworkin could point out, with some validity, that the capabilities approach is hazy on what exactly equality of capabilities would involve, in terms of transfer calculations. He might, further, contend that an average hypothetical insurance model provides the best approximation. However, in this case, it would not be equality of capability that collapsed into equality of resources, but rather that the capability approach provided equality of resources with a theoretical backdrop of which it was much in need. (See also Footnote 50)

<sup>35</sup> Dworkin 2000, 303

of capabilities collapses into equality of resources, this claim would have to be phrased, more strongly, in terms of rankings of *capabilities*.

This would not hold for a version of the capabilities approach (such as the one I will offer in Section IV), which holds that the appropriate metric for distributive justice is capability to achieve *each* of the functionings on something like Nussbaum's list, not just those whom each individual happens to value.

Note, moreover, that if it is enough to satisfy a capabilities-based model of distributive justice that individuals are rendered equal in their capacity to pick and choose across central capabilities, the corresponding view at the wider level must be, not that individual human flourishing requires *all* the capabilities, but rather that it requires only such capabilities as the individual would choose to have. Thus, behind Dworkin's *resources*-based account of redistributive justice must be an implicit reliance (which he sees the capabilities approach as sharing) on preference or choice satisfaction as the underlying element of a flourishing human life.<sup>36</sup> This is the Preferences Accusation, and can thus be rebuffed by solving the PPP.

### III. The Dworkinian complaint 2: welfare & paternalism

Dworkin's other reading of Sen is 'as suggesting that people should be made as nearly equal as possible in their capacity to realise the "complex" achievements of happiness, self-respect, and a significant role in the life of the community'.<sup>37</sup>

On this understanding, Dworkin claims, equality of capabilities collapses not only into equality of welfare but into 'a particularly chilling form'<sup>38</sup> of it. Of course, he explains, 'it is good when people are happy, think well of themselves, and are thought well of by others. The idea that people should be *equal* in their capacities to achieve these desirable states of affairs, however, is barely coherent and certainly bizarre – why should that be *good*? – and the idea that government should take steps to bring about that equality... is frightening.'<sup>39</sup>

---

<sup>36</sup> This fits well with his saying 'it is good when people are happy...' but sits less well with the end of this sentence: '...and are thought well of by others.' (Dworkin 2000, 302) This last suggests a possible determinant of the "goodness" of a human life which is not entirely dependent on the individual's own mental state.

<sup>37</sup> Dworkin 2000, 301

<sup>38</sup> Ibid., 301-2

<sup>39</sup> Ibid., 302

To elaborate, Dworkin distinguishes between two aims. Firstly, the (legitimate) one of redistributing resources the lack of which renders it impossible to achieve such ends as happiness, self-respect, and a decent role in community life. Secondly, the (he thinks, illegitimate) one of making people 'equal in their overall capacity to achieve these goals, whatever ambitions, projects, tastes, dispositions, convictions, and attitudes they might have, which is the false goal of equal welfare, or well-being'.<sup>40</sup>

This is ambiguous, as can be brought out by asking why it is necessarily a bad thing for equality of capabilities to turn out to be a form of equality of welfare. Equality of welfare, understood at its broadest as equality of *well-being*, has considerable intuitive appeal. To reject equality of capability on these grounds, Dworkin would have also to show that it amounted either to a version of equality of welfare which falls under his own, expensive taste-related criticisms,<sup>41</sup> or to one which is counterintuitive for some other reason.

It is unclear which of these paths he is taking, but let us begin by considering the former. Several of his remarks suggest this. He points out that '[p]eople vary in their capability for "happiness" for a thousand reasons, including their wealth, their personality, their ambitions, their sensitivity to the suffering of others, and their attitudes towards Milton's rival muses. Molière's misanthrope lacks the capacity for happiness of Voltaire's Pangloss.' He stresses that what makes it impossible for many to achieve 'happiness, self-respect, and a decent role in community life' is not only a lack of external resources, but also, often, a lack of internal ones.<sup>42</sup> This appears to echo his own earlier attack on equality of welfare, measured subjectively.

Even within this line of criticism, there is scope for interpretation. Dworkin could be assuming that the capabilities approach would pander to individuals' own perceptions of their quality of life, to the exclusion of all more factual criteria, effectively collapsing into exactly the kind of equality of welfare which he has already, famously, attacked. Alternatively, he could be objecting to the inclusion of *any* subjective criterion in the metric for equality.

---

<sup>40</sup> Ibid., 302

<sup>41</sup> Ibid., 44-64

<sup>42</sup> Ibid., 302

If the former, he misunderstands the capabilities approach. Expensive taste objections to equality of welfare point to the intuitively repugnant possibility that (say) Tiny Tim could end up with fewer resources than Scrooge because, despite his handicaps and other physical and nutritional disadvantages, he is happier and has more self respect.<sup>43</sup> However, appeal to such counterexamples relies on stressing Sen's mention of "happiness" as a capability (one which Nussbaum does not include) at the expense of others, including Sen's own "elementary" (more objective) functionings, such as good health and nutrition.<sup>44</sup> Far from denying Tiny Tim compensation, the capabilities account is designed expressly to take account of such objective criteria as his nutritional or medical needs.

Moreover, to accuse equality of capabilities of exclusive reliance on subjective criteria, Dworkin must, again, take the approach at the wider level (as an account of what is important for human flourishing) to rest entirely on how satisfied individuals themselves are with their lives. This is the Preferences Accusation, which will be denied in the following section.

Given the inaccuracy of such a reading of the capabilities approach, a more generous interpretation of Dworkin is Williams' revised version. However, this, as we shall see, is no longer a straightforward accusation of "collapse". Williams suggests:

[T]he resourceist should not argue that if the capability view departs from equality of resources, it becomes equivalent to equality of welfare. Instead, he should allow that a capability view can differentiate itself from equality of resources by focusing on complex capabilities yet also retain the nonderivative concern for elementary capabilities which distinguishes it from welfare. He should nonetheless still insist that, to the extent that such a view departs from equality of resources, it does so only by embracing a concern with welfare which, even if not exclusive, renders it vulnerable to antiwelfarist objections.<sup>45</sup>

Williams states, not implausibly, that 'Dworkin's reference to "collapse" is best understood in this sense'.<sup>46</sup> But the accusation has changed: there is an important difference between the claim that equality of capabilities must either be no different from equality of welfare, or no different from equality of

---

<sup>43</sup> Dworkin's own example (2000, 59-62)

<sup>44</sup> Sen 1999, 39-40

<sup>45</sup> Williams 2002, 27

<sup>46</sup> *Ibid.*, 28

resources, and the claim that it must either be no different from equality of resources, or contain a subjective element which renders it vulnerable to the same attacks as can be made on equality of welfare.

In the latter case, the capability approach would remain distinct from either of its two rivals, with the success of the Dworkinian attack resting not straightforwardly on a reference back to his earlier expensive taste-based rejection of equality of welfare, but rather on the claim that his earlier (or other) arguments can justify rejecting *any* metric for distributional equality which incorporates *any* subjective welfarist element. An account is owed of why this should be. Moreover, the following points suggest that Dworkin's own counterexamples cannot so easily be applied.

Firstly, the *inclusion* of such criteria does not mean considering them *exclusively*. Tiny Tim's state of mind might mean that he does not, as some others might, have a claim on the tax system for treatment for depression. However, his health problems and his lack of adequate nutrition give him a claim to extra resources, or social facilities, to boost him in terms of those capabilities. Indeed, one of the motivating factors in developing the capabilities approach, as an alternative to a welfarist theory, is adaptation to poor lifestyle conditions or mistreatment. Expensive taste criticisms, the other side of the same coin, are similarly defused by the inclusion of objective criteria to balance out subjective claims.

The Dworkian objection might be reworked, pointing to the intuitive dis appeal of someone's having *any* redistributive claim, even one ultimately balanced out, on the grounds of voluntarily-acquired expensive tastes. Surely, we feel, Louis' consciously cultivated taste for plovers' eggs should not give him any claim to compensation.<sup>47</sup>

However, it is far from clear that the capabilities approach would be committed to compensating him. Much of the force of Dworkin's expensive taste arguments can be removed by shifting the goal from equality of actual welfare to equality of *opportunity* for welfare. They are similarly defused by the distinction between functionings and capabilities. Such criticisms rest on the false idea that equality of capabilities would seek to give people (say) equal self-respect rather

---

<sup>47</sup> Dworkin's example (2000, 49-52)

than what it would actually target: the *capability* to achieve self respect, or, as Nussbaum puts it, ‘the social bases of self-respect and non-humiliation’.<sup>48</sup>

Moreover, Dworkin’s attack rests heavily on Sen’s reference to “happiness”. This seems to operate at a different level to the other suggested functionings, and does not feature on Nussbaum’s list. Instead, more specifically, she includes the ability ‘to have pleasurable experiences’ within Senses, Imagination and Thought, and lists Play (‘Being able to laugh, to play, to enjoy recreational activities’) and Emotions as central capabilities.<sup>49</sup>

There is also an important distinction between central capabilities (broad-based capability “categories”, such as “Life” or “Affiliation”) and subcategory capabilities. Thus, for example, even if distributive claims were based on *functioning* shortfalls, it is unlikely that Louis would have any claim. Although his pleasure might be reduced by the fact that he cannot indulge his tastes as frequently, he has, we can assume, made other gains within the category of Senses, Imagination and Thought. He has searched for his own meaning of life in his own way (he has, as Dworkin himself sets up the example, cultivated this taste precisely because he sees it as making his life more worthwhile<sup>50</sup>) and presumably thought, reasoned and imagined in the process.

Thus, on one (entirely) subjectivist reading of the capabilities approach, Dworkin’s argument relies again on the Preferences Accusation. On another (partially) subjectivist reading, he does not demonstrate collapse into equality of welfare, or even into an account sufficiently welfarist to fail by appeal to the

---

<sup>48</sup> Nussbaum 2000b, 79-80

<sup>49</sup> Ibid., 80

<sup>50</sup> A related line of attack focuses on another of Dworkin’s counterexamples to welfarism: “bottomless pit” problems which arise when an individual’s physical condition is so dire that they could never be made “equal” with regard to some subjective criterion. (2000, 60-1) However, this applies not just to subjective criteria but also to more objective ones such as health. Some individuals will have such bad handicaps that even given almost all of society’s resources they would never be equal to others in a particular capability. This could be a stronger case than those considered above for a shift to a hypothetical insurance model, as the practical way round these problems. However, a) it will be clear from the previous and next sections that the hypothetical insurance calculations would have to operate *within* each central capability, the motivating idea still being to get as close as possible as equality in each of these, rather than to provide ability to choose and make trade-offs across them; b) this would not touch the approach at the broader level, as a model of human flourishing – it would limit the criticism to the sphere of distributive justice; c) the presence of some exceptional problem cases does not seem adequate grounds for giving up altogether on distribution according to capabilities, especially when Dworkin’s alternative would permit significant discrepancies in capabilities in a wider range of cases than just these extreme scenarios.

same expensive taste-related counterarguments. We are left, then, with the alternative reading of Dworkin highlighted at the start of the section.

Rather than attacking the capabilities approach for making too much space for individual tastes and ambitions, he could be accusing it of collapsing into a version of equality of welfare which is repugnant for exactly the opposite reason – that is, which does not leave *enough* room for individual taste.

There is, again, textual evidence for this. Although he ignores Sen's more objective criteria (good health, nutrition, avoiding escapable morbidity and so on<sup>51</sup>) and cites both "happiness" and "self-respect" (another subjective criterion) as among the Senian 'important goods' for human beings, he also acknowledges the more objective functioning of "taking part in the life of the community". Being 'thought well of by others' is a measure of quality of life *external* to the individual.<sup>52</sup>

Thus, there is a plausible reading on which Dworkin is accusing the capabilities approach of drawing up a set of criteria for quality of life and imposing this on individuals regardless of their individual 'ambitions, projects, tastes, dispositions, convictions, and attitudes'.<sup>53</sup> In this case, his references to the "frighteningness" of political interference, and the "chilling" form of equality of welfare make more sense. They draw on the idea that it should be *up to* the individuals whether, and to what extent, their quality of life depends on such criteria as (say) involvement in the life of the community.

This is an important objection. It is, moreover, one which applies at the broader level. But the PPP already takes it into account, as the Paternalism Accusation. Thus, again, defeating Dworkin on this score becomes one side of solving the PPP.

#### **IV. More than preference satisfaction**

I am left with the twofold task of solving the PPP. The Preferences Accusation, recall, claims to reduce the capabilities approach to a solely preference-based account of human flourishing. The first point to press in refuting this is, as has

---

<sup>51</sup> Sen 1992, 39

<sup>52</sup> Ibid., 302

<sup>53</sup> Ibid., 302



frequently been pointed out, the difference between functionings and capabilities. While space is left for individual freedom in choosing whether or not to exercise the functionings, that does not mean that an approach which identifies capabilities as prerequisites for human flourishing must collapse into preference satisfaction.

Rather, the question is whether scope is allowed for individuals to choose between the central *capabilities*, as expressed by something like the categories on Nussbaum's list. I will simply bite the bullet. That is, I will take the capabilities approach as saying that the functional capabilities are *all* necessary elements of a flourishing human life, regardless of whether each individual thinks so. This implies that it is not in an individual's ATC interest to give up, permanently, any one of them.<sup>54</sup>

There are three things, however, that I want to be careful *not* to say. Firstly, to reiterate, there is no suggestion that it is not up to the individual whether to exercise the relevant *functioning*. Consider examples such as the rich faster, the monk adopting an ascetic life style, or George Orwell down and out in Paris and London.<sup>55</sup>

Nor, secondly, am I claiming that there should be no scope for an individual to put certain capabilities *at risk* in order to function more in others. This would be ridiculous: some degree of risk is an unavoidable part of life. Moreover, it is strongly intuitive that, when at least a certain level of risk is balanced against an autonomously chosen goal (one which, perhaps, forms an important element of her rational plan of life<sup>56</sup>), it must be up to the individual to say what is in her interest, *all things considered*.<sup>57</sup> To deny this would be to deny, for example, that it can contribute to flourishing to take part in many sports.

---

<sup>54</sup> As I will clarify in the final section, this claim can be understood in two ways. However, for the moment it is sufficient to lay it out like this. It is worth re-stressing that this chapter aims to lay out a version of the capabilities approach which can escape the PPP, *not* to defend such a version as exactly in line with either Sen or Nussbaum's accounts. However, there is support for this reading in Nussbaum. For example: 'The "capabilities approach", as I conceive it, claims that a life which lacks *any one* of these [central functional] capabilities, no matter what else it has, will fall short of being a good human life.' (1998, 320)

<sup>55</sup> These are Wolff & De-Shalit's examples, with the rich faster borrowed from Sen (Wolff & De-Shalit 2007, 64-5; Sen 1999, 75)

<sup>56</sup> Recall Nussbaum's Category 6: Practical Reason, which includes being able 'to engage in critical reflection about the planning of one's life'. (Nussbaum 2000b, 79)

<sup>57</sup> I will come back in the next section to the question of autonomous choice.

However, there is an important difference between this and a deliberate decision irrevocably to give up a capability (or to expose oneself to a very high risk of doing so). Of course, I cannot identify precisely at which point on the risk scale this kicks in. It is enough to say that, on this version of the capabilities approach, it is contrary to a person's having a fully flourishing life for her to give up, or to almost certainly give up, voluntarily or otherwise, all future capability for one of (something like) Nussbaum's functionings. Moreover, there are clear grounds for concern if an institutional framework enables individuals to exercise certain capabilities only by putting others in serious jeopardy.<sup>58</sup>

Thirdly, it is perfectly compatible with this that individuals might choose, rationally, to give up one of these capabilities for some external end – most plausibly the life or well-being of some other person or people. To say that all these capabilities are necessary constituents of a flourishing life is not to say anything about when a person might be justified in sacrificing (or might even be morally bound to sacrifice) some part of their own flourishing.

The Preferences Accusation fails against this version of the capabilities approach. The claim that it is a prerequisite for a fully flourishing life that an individual has *all* the central capabilities draws a clear distinction between this and any model which identifies a flourishing life with one in which the individual's own preferences are satisfied.

Correspondingly, Dworkin's attempt to reduce equality of capability, as a political goal, to equality of resources, is mistaken. As Section II demonstrated, the alleged collapse depends on the assumption that equality of capabilities would not require equality in *each* central capability, but only a 'nearly equal'<sup>59</sup> background, in terms of resources, against which each individual could choose such capabilities as she valued. Assuming that the aim of distributive justice is to provide all individuals with an effective opportunity to live a fully flourishing life, the version of the capabilities approach just given would not be satisfied with this.

A final note, before addressing the Paternalism Accusation. I have defeated the Preferences Accusation as it stands. However, it might be reformulated as the

---

<sup>58</sup> Consider Sen's example of Kader Mia, who is obliged to risk his life in order to work. (1999, 8)

<sup>59</sup> Dworkin 2000, 303

somewhat different claim: that the capabilities approach collapses into so-called *ideal* or *rational* preference satisfaction. Call this the Ideal Preferences Accusation.

On an ideal preference satisfaction account, a flourishing life is characterised by the satisfaction of preferences adjusted in certain ways, or the preferences a person would have in certain ideal conditions. For example, on Richard Brandt's account:

[A] person's desire, aversion, or pleasure [is] 'rational' if it would survive or be produced by careful 'cognitive' psychotherapy for that person. I shall call a desire 'irrational' if it cannot survive compatibly with clear and repeated judgements about established facts.<sup>60</sup>

He points to four primary types of "mistaken" desires, aversions, or pleasures: dependence on false beliefs; artificial desire-arousal in cultural transmission; generalisation from untypical examples; and exaggerated valences produced by early deprivation.<sup>61</sup> Cognitive psychotherapy is supposed to correct for these.

It should be clear that the capabilities approach, as laid out here, does not collapse into this kind of ideal preference satisfaction. There is no reason to assume that all individuals, taken from current circumstances, would *prefer* or *choose* the capabilities even after the kind of cognitive psychotherapy advocated by Brandt.

Moreover, even if the two did happen to coincide, for the alleged collapse into ideal preference satisfaction to take place, the capabilities would have to be mere *means* to ensure that preferences were rational, with the satisfaction of rational or ideal preferences regarded as, ultimately, constituent of human flourishing. But even on a "practical reasoning"-based account of the capabilities, such as the model favoured by Sabina Alkire, the line of argument goes the other way.<sup>62</sup> Reason (like Nussbaum's deep and broad consensus<sup>63</sup>) is appealed to as the best way of "getting at" the capabilities essential to flourishing. This is not the same

---

<sup>60</sup> Brandt 1979, 113

<sup>61</sup> See Brandt 1979, 115-26

<sup>62</sup> Alkire, inspired by Finnis, defends appeal to reflection on the question "why do I (or others) do what I (or they) do?" to identify 'a discrete heterogeneous set of most basic and simple reasons for acting which reflect the complete range of kinds of valuable human states and actions (the complete range of functionings)'. (2002, 46, referring to John Finnis: *Natural Law and Natural Rights*, Oxford University Press, 1980)

<sup>63</sup> Nussbaum 2000a, 120

as an approach which assigns ultimate importance to the satisfaction of rational preferences, and gives the capabilities the instrumental role of ensuring that preferences are formed rationally.

Anticipating later sections, however, an understanding of ideal preferences can be envisaged such that the two accounts might well converge. Suppose ideal preferences were defined as those preferences that *would be* formed under something like the Razian conditions for autonomous choice.<sup>64</sup> It is plausible that the capabilities would turn out to correspond to what (at least most) individuals would prefer in ideal conditions. However, this is not a problem for the capabilities account. Indeed, as the discussion of Section I should have made clear, it is an advantage.<sup>65</sup>

Again, the direction of the argument need not be the same. The account could agree that the only way, ultimately, of getting at what *is* necessary to human flourishing is through such ideal preferences. But it would *not* thereby have to concede that what is ultimately valuable is the satisfaction of ideal preferences, *in itself*. Even were a capabilities theorist to concede that, in the end, little turns on such a distinction, this hypothetical understanding of ideal preferences goes a long way beyond undiluted preference satisfaction, or indeed Brandt-style accounts which take existing preferences and somehow “correct” them.

Thus, only on a very specific reading of “ideal” could enjoyment of the capabilities plausibly correspond with ideal preference satisfaction. Such an model would be so far distinct from actual preference-based accounts as to continue to evade the Preferences Accusation. However, this does raise some specific questions regarding freedom to choose between capabilities *within* the circumstances of autonomous choice. I will return to this in Section VI.

---

<sup>64</sup> Absence of coercion and manipulation; appropriate mental faculties; and adequate choices. (Raz, 1986, 372-2). These will be discussed further in Section V.

<sup>65</sup> Nussbaum, while rejecting preference or desire-based accounts of the good, stresses: ‘nonetheless, it is quite important to political justification that there should be a good measure of convergence between a substantive-good approach and an intelligently designed informed-desire approach’. (2000a, 117-8)

## V. Paternalism: acceptable and unacceptable

Turning to the other side of the PPP, the Paternalism Accusation says, broadly, that adult human beings are the best judge of what is in their own interest, and it is unacceptable to impose on them any external conception of the good.

One narrower version of the criticism focuses on the unattractiveness of the idea that *philosophers* consider themselves in a position to speak authoritatively on what is best for other people. It is intuitively unappealing that a theorist should, with something like the intellectual snobbery implicit in John Stuart Mill's ranking of pleasures, pass judgement on what is good for the rest of the humanity.<sup>66</sup>

This is a forceful point. At first sight, Nussbaum's account could, indeed, appear overweighed to the philosophical viewpoint, stemming as it does from ideas in Aristotle and Marx. However, as discussed in Section I, her later reformulation of her position defends the list of functional capabilities as derived through a procedure which is 'neither ahistorical nor a priori. It is the attempt to summarise the empirical findings of a broad and ongoing cross-cultural inquiry. As such, it is open ended and humble: it can always be contested and remade.'<sup>67</sup> She draws extensively on work with women in developing countries, notably India.

Moreover, Jonathan Wolff and Avner De-Shalit have since attempted to validate their slightly modified version of the list 'by means of a cross-check by more empirical forms of enquiry: consultation with people from a wide variety of walks of life, analysis of surveys, and the like'.<sup>68</sup> This goes some way towards rebutting this narrower version of the paternalism objection, although the range of opinion canvassed by Wolff and De-Shalit was relatively limited, and

---

<sup>66</sup> Mill 1861, 136-42

<sup>67</sup> Nussbaum 1998, 317

<sup>68</sup> Wolff & De-Shalit 2007, 43-4. See Footnote 22 for their additions. This evaluation process had four parts: (1) Reflection upon the list 'in a traditional way' (43) through consideration of their own experience of the disadvantaged and discussion with other political philosophers and those in social policy, as well as consideration of existing empirical studies. (2) Proposal of further categories. (3) Testing of the expanded list through a total of 88 formal interviews with welfare services providers and disadvantaged individuals in Israel and in England. (4) Asking interviewees whether they felt that any categories were missing. The response was general affirmation but with some categories (notably (8), Other Species) considered less central, and one addition suggested: 'the functioning of being able to communicate, including being able to speak the local language, or being verbally independent'. (60)

restricted to the UK and Israel. A more thorough endorsement would require research across more diverse cultures.

However, such consensus on this particular list does not refute the more general paternalist critique. As Nussbaum summarises it, the accusation is that

when we use a set of universal norms as benchmarks for the world's various societies, telling people what is good for them, we show too little respect for people's freedom as agents... People are the best judge of what is good for them, and if we prevent people from acting on their own choices, we treat them like children.<sup>69</sup>

One immediate response is to stress, again, as Nussbaum does, that her list is one of *capabilities*.<sup>70</sup> What matters is that each individual human being has the *option* of functioning in each of the categories. She thus steers clear of any dangerous suggestion that individuals should be "forced to be free".

However, the charge of paternalism can still be pressed. Is it not paternalistic to judge a person's quality of life according to a set of *capabilities* which she herself might not choose to prioritise, or to assign interests to her which she does not embrace as her goals? This section will argue that the capabilities approach is not *unacceptably* paternalistic.

Begin by considering what exactly it is that we object to about paternalism. The following might be taken as a first draft.

#### *First Paternalism Assertion*

It is unacceptably paternalistic to judge the success or otherwise of another human being's life according to a thick conception of the good that does not correspond to her own conception of the good.

The term "thick" is used here in contrast to a Rawlsian liberal "thin theory" of the good.<sup>71</sup> For Rawls, famously, the right is prior to the good. Individuals have their own comprehensive conception of the good and rational plan of life, and freedom to pursue these is central. However, parties behind his veil of ignorance,

---

<sup>69</sup> Nussbaum 2000b, 51

<sup>70</sup> The relation between capabilities and functions is, she stresses, "crucial to... [addressing] concerns about paternalism and pluralism... Where adult citizens are concerned, *capability, not functioning, is the appropriate political goal.*" (Nussbaum 2000b, 86-7 [her emphasis])

<sup>71</sup> Nussbaum 1990, 210.

ignorant of their own comprehensive views, agree on the primary goods (rights, liberties, opportunities, income and wealth, the social bases of self respect<sup>72</sup>) as a proxy or “thin” conception of the good. They then choose principles on the assumption that more primary goods is better than less. A thick theory of the good takes a more comprehensive stance on what is needed for a flourishing human life.

As it stands, the Paternalism Assertion can be questioned. Indeed, its force depends on two factors: the *preciseness* of the conception of the good imposed; and the conditions under which the individual’s own conception of the good is held or adopted.

Consider the latter first. The thin liberal objection to something like Nussbaum’s list is that it overrides individual choice. This rests on the idea that what matters most for humans is the ability to decide how to live their own lives. Thus only goals and interests acknowledged by the individuals themselves should be allowed to “count” in determining whether they are fully flourishing, or what is in their interest. Because the capabilities approach doesn’t limit itself in this way, the suggestion is, it is unacceptably paternalistic.

However, the capabilities theorist has a reply to this which not only serves as a response to the Paternalism Accusation, but also indicates the flaws in the kind of preference satisfaction account of human flourishing which Dworkin, as well as Rawls, appears to assume to be a good thing.

The force of the thin liberal objection – the supposed intuitive *unacceptability* of a capabilities model – hinges on a misconception of what is required for individuals to be able to decide, in some genuine sense, how to live their lives. *Meaningful* choice requires more than negative freedom to choose against whatever background of living conditions, family, educational opportunities, and so on, a person happens to have.

Nussbaum offers examples of women in India with established entrenched “preferences” that reflect the ill-treatment to which they have become accustomed.<sup>73</sup> Even in a supposedly advanced Western society, a young adult who has spent his childhood watching TV in a small flat might “choose” to stay in and watch *Big Brother* every evening rather than go out to the theatre, play a

---

<sup>72</sup> Rawls 1971, 54

<sup>73</sup> Nussbaum 2000b, 114

sport, or attend a night school to get the chance to go to university. A woman who has been beaten by her father in her childhood, might “choose” a similarly abusive husband, and then “choose” not to leave him. But this is not the kind of “free” choice on which we should be placing so much value.

It is important to stress once again that the capabilities theorist does not attempt, Mill-style, to classify certain activities as more valuable than others.<sup>74</sup> The claim is simply that assigning absolute priority to an individual’s chosen conception of the good or plan or life loses its appeal if that conception has been imposed on her by the circumstances in which she lives, if she has never had the opportunity to reflect on it, to express freely her own thoughts, and to discuss and develop her own views within a wider informational context.

It would be better for the young man (his life could be said to be going better) if he were deciding against a background of a childhood with the opportunity to explore various different ways of spending his time, and the basic education required to be able to appreciate them. Similarly, the woman’s quality of life would be higher if she could (various other assumptions being made) get up the courage to leave her husband, or not have married him in the first place. It might well, thus, be in her interest to do so even if she persuades herself that it is not. We explain this intuition by appealing to the idea that she is rendered unable, by circumstances, to value herself as she ought, exercise autonomy, and escape from the situation.

With such examples in mind, the hypothetical objector’s claim, that what “matters most” to individuals is being able to choose how to live their own lives, is more plausible understood as a claim about autonomy, *not* about negative freedom. The appeal lies in what Raz identifies as ‘[t]he ruling idea behind the idea of personal autonomy... that people should make their own lives’.<sup>75</sup> Given this, the capability approach has a twofold defence to the thin liberal objection. Firstly, as Nussbaum herself has stressed, it does not exclude considerations of a person’s autonomously or genuinely chosen plan of life or conception of the good. Secondly, unlike a “thin” preference-based account, it establishes the background criteria likely to ensure that an individual *is* deciding autonomously.

---

<sup>74</sup> See Mill 1861, 137-42

<sup>75</sup> Raz 1986, 369



Begin with the first point. Far from under-weighting the individual's own choices, Nussbaum gives central importance to an individual's *genuinely espoused* conception of the good and rational plan of life.<sup>76</sup> This comes out most strongly in capability (6), Practical Reason, which specifies '[b]eing able to form a conception of the good and to engage in critical reflection about the planning of one's own life',<sup>77</sup> and is part of a more general emphasis on meaningful choice and human *individuality* as necessary components of a flourishing life. Functional capabilities (3), Bodily Integrity, (4), which includes being able 'to search for meaning of life in one's own way',<sup>78</sup> (7B) and (10), as well as (6), make this particularly clear.

This should come as no surprise. Recalling the quote from Sen which opened Section I, the fundamental aim of the capabilities approach is to ensure that individuals have 'the *substantive* freedoms – the capabilities – to choose a life [they have] reason to value'.<sup>79</sup>

To make the second point, we need a clearer specification of what is meant by an autonomous, genuine, or meaningful choice. Such an account has been provided by Joseph Raz, according to whom an autonomous person's well-being 'consists in the successful pursuit of self-chosen goals and relationships'.<sup>80</sup> He identifies three 'conditions of autonomy', only one of which is the absence of 'coercion and manipulation by others' standardly associated with negative freedom. The others are: 'appropriate mental faculties... [including] minimum rationality, the ability to comprehend the means required to realise [one's] goals, the mental faculties necessary to plan actions etc.'; and 'an adequate range of choices'. Within this last, Raz stresses the need for 'options with long term pervasive consequences as well as short term options of little consequence, and a fair spread in between'.<sup>81</sup> Variety, not just number, is essential.<sup>82</sup>

Negative freedom, against whatever historical and background circumstances happen to hold, is no guarantee that these conditions will obtain. But there is

---

<sup>76</sup> A point reinforced by the fact that communitarians have accused her of being *too* individualistic. (Robeyns responds to such accusations (2005, 107-9))

<sup>77</sup> Nussbaum 2000b, 79

<sup>78</sup> *Ibid.*, 79

<sup>79</sup> Sen 1999, 74 [emphasis added]

<sup>80</sup> Raz 1986, 370

<sup>81</sup> *Ibid.*, 372-4

<sup>82</sup> *Ibid.*, 375-7

reason to believe that the capabilities, on something like Nussbaum's list, will do rather better. Not only does Bodily Integrity feature as a central capability, but adequate education is stressed under Senses, Imagination and Thought, critical reflection is stressed under Practical Reason, and the very range of capabilities and sub capabilities indicates a range of long and short term choices in line with Raz's condition.<sup>83</sup>

Having thus offered one strong line of defence against the Paternalism Accusation, by identifying a way in which the First Paternalism Assertion fails to get at what is really unacceptable about paternalism, I will turn to the first of the two factors to be considered. What sort of conception of the good is being "imposed" on individuals by the capabilities approach, in the sense that it is being used to judge how well or badly their lives are going? This is not what might be called a "thick, *precise* conception" of the human good, which lays out exactly which way of life should be adopted, and provides detailed rules (along the lines of a religious code) for living by it.

Rather, Nussbaum's is a *vague* conception. She outlines functionings at a very general level, based on intuitive, uncontroversial and, importantly, open-ended ideas about what human flourishing requires – ideas on which there can be broad consensus at a deep level.<sup>84</sup> Value is assigned to the ability to: form and live by

---

<sup>83</sup> This very defence opens up another line of questioning: is it possible to have substantive freedom, or consistently to be making one's choices autonomously, without some of the central capabilities? If so, why should the *full* list be taken as necessary for human flourishing, rather than just the Razian conditions? In response, note firstly and obviously that if someone is, for natural reasons, unable to enjoy a central capability (say, health), it remains important to seek to ensure that, *so far as is possible*, their choices are made autonomously. Secondly, when it comes to certain key decisions, such as espousing a conception of the good, it is at least arguable that someone who didn't enjoy all the central capabilities would have an inadequate range of options. It is to precisely this idea that Nussbaum appeals in defending her model: 'If one cares about people's powers to choose a conception of the good, then one must care about the rest of the form of life that supports those powers, including its material conditions.' (1998, 324) However, Joshua Cohen argues that Sen cannot rest his appeal to capabilities on the importance of substantive freedom *alone*. Capabilities, Cohen says, are a wider concept, and he appeals to examples such as improvements in drinking water quality as improving capability to function without making an individual more *free*. (1995, 278-80) Moreover, Nussbaum uses the distinction between capabilities and functionings to separate her own view from that of liberal perfectionists (of whom Raz is one), although she does concede that she is 'in a sense perfectionist' (2000a, 129) about practical reason. Her focus on respect for choice should, she specifies, be understood 'not in terms of a comprehensive liberal ideal of autonomy but in terms of an idea of respect for the diversity of persons and their comprehensive conceptions'. (ibid.) She would thus deny that autonomous choice is the *only* factor in a flourishing life. However, it is enough, to defeat the Paternalism Accusation that the list of capabilities can be defended as being *better* able to allow individuals to make meaningful, or genuinely "free" choices, than thin liberal negative freedom alone.

<sup>84</sup> Nussbaum 1990, 219; 1998, 318; 2000b, 77

*some* conception of the good; make and carry out *a* plan for one's own life; produce 'self-expressive works and events of one's own choice, religious, literary, musical, and so forth';<sup>85</sup> and engage in 'various forms of social interaction'.<sup>86</sup> Opportunities are what matters; no comprehensive conception is being imposed on anyone.

The above arguments reveal the First Paternalism Assertion as inadequate. To identify more plausibly what is unacceptable, it needs to be rewritten along the following lines.

*Strong Paternalism Assertion*

It is unacceptably paternalistic to judge the success or otherwise of another human being's life according to a thick, precise conception of the good which does not correspond to her own considered conception of the good, formed within the conditions for autonomous choice.

This may be true but it fails, as we have seen, to get a grip on the capabilities approach. There may be elements of paternalism in the approach (such elements, one might suggest, are unfortunate but necessary in a world rife with entrenched preferences and non-autonomous choice) but these do not render it unacceptably paternalistic. As Nussbaum herself puts it:

We dislike paternalism, insofar as we do, because there is something else that we like, namely each person's liberty of choice in fundamental matters. It is fully consistent to reject some forms of paternalism while supporting those that underwrite these central values, on an equal basis.<sup>87</sup>

The capabilities approach has been defended against the Paternalism Accusation. This completes the solution of the PPP. However, there remains another potential objection, suggested by the earlier consideration of ideal preference satisfaction, and by close attention to the Strong Paternalism Assertion.

---

<sup>85</sup> Nussbaum 2000b, 79

<sup>86</sup> Ibid., 79

<sup>87</sup> Ibid., 53. Raz, in his perfectionist critique of liberal neutrality, makes a similar point, drawing a distinction between comprehensive and narrow neutrality. (1986, 123-4)

## VI. A final objection

According to the version of the capabilities approach presented here, it is not in someone's interest (would be detrimental to their flourishing) to give up, permanently (or expose to very serious risk) one of the functional capabilities on something like Nussbaum's list. I have defended this as not unacceptably paternalistic. In so doing, I have pointed to the Razian conditions for autonomous or meaningful choice. Suppose, however, that an individual who had hitherto had all the capabilities, and so satisfied the Razian conditions, decided permanently to give one of them up. (Not, moreover, as a sacrifice in pursuit of some external goal, as discussed in Section IV.) What would the capabilities approach say?<sup>88</sup>

Note, firstly, that such cases are likely to be rare. Giving up a capability is very different to giving up a functioning. Moreover, there is a difference between giving up a sub-category capability, on something like Nussbaum's list (say, by being voluntarily sterilised), and giving up (say) one's entire capability for health, or affiliation. It is therefore worth reiterating that reference to "capabilities" should, in this discussion, be taken as reference to the *categories* on something like Nussbaum's list.

The question, then, is whether the capabilities approach must deny the following claim.

### *Permissibility Claim (a)*

An adult human can have a fully flourishing life despite giving up a central capability, provided that she has enjoyed all the capabilities (and so been in the circumstances of autonomous choice) up to the point.

---

<sup>88</sup> This also brings us back to the point left open by the discussion of the Ideal Preferences Accusation at the close of Section IV. If the capabilities did indeed correspond with ideal preferences, defined as the satisfaction of preferences made within the conditions of autonomous choice, they could do so in two ways. Is the satisfaction of *each individual's* ideal preferences, so defined, supposed to get at what is necessary for that individual's flourishing? Or is the idea that such ideal preferences, *in general*, offer a guide to what is generally necessary for human flourishing? In the former case, there is less distinction between the capabilities approach and an ideal preference satisfaction account. In the latter, the distinction is maintained, but arguably at the cost of imposing a view external to each individual, in a way which is unacceptably paternalistic. This turns on our exact understanding of the Strong Paternalism Assertion, and so is addressed by the arguments of this section.

Insofar as it relates to the use of the capabilities approach in the rest of this thesis, the claim could be rewritten as:

*Permissibility Claim (b)*

It can be in the ATC interest of an adult human permanently to give up a central capability, provided that she has enjoyed all the capabilities (and so been in the circumstances of autonomous choice) up to the point.<sup>89</sup>

My contention (perhaps a frustrating one for the reader) is that the capabilities approach could go either way without falling foul of the PPP. There are two possible versions. The first (call it the Non-Permissive Version) would deny the Permissibility Claims. This need not render the approach unacceptably paternalistic, provided we take a certain reading of the Strong Paternalism Assertion.

To clarify, the Strong Paternalism Assertion as it stands suggests that both of its conditions (preciseness of the imposed conception of the good, and formation of the individual's own conception of the good under the circumstances of autonomous choice) must be satisfied to avoid unacceptable paternalism. However, it could be rephrased, according to the outcome of the question (which the arguments here have not settled) whether paternalism still counts as unacceptable if only the second condition is met. That is, if the conception of the good imposed is vague, but is nonetheless at odds with what an individual has formed as her conception of the good under the conditions of autonomous choice.

As the Strong Paternalism Assertion stands, the Non-Permissive Version would not be unacceptably paternalistic because nothing has been said to threaten the defence, above, of a Nussbaumian theory of the good as a *vague* thick conception. If, however, the connection between the two conditions is

---

<sup>89</sup> As already stressed, the understanding of the capabilities account offered here does not pretend to be an exact reading of Nussbaum's own view. In any case, it is hard to conclude from her work what the approach should say in this scenario, given that she considers only the possibility of an adult's signing away a major capability, permanently, 'apparently without coercion', in which case she thinks that '[f]requently, though certainly not always, we will judge that interference is justified to protect the capability'. (Nussbaum 2000b, 93) Not only does she not address the question under the more stringent Razian conditions, but her focus here is on what *political* action is justified, not on the more basic point of whether the person is acting against his or her interest or flourishing.

understood as an “or” rather than an “and”, the Non-Permissive Version could be harder to maintain. (Although, to reiterate, there is little evidence that such cases would arise. In the real world, when such choices are made, it is very rarely against such backgrounds.)

This need not trouble the capabilities approach, which could equally well adopt the Permissive Version. This stops short of denying the Permissibility Claims, requiring for flourishing only that an individual have all the capabilities at least to a certain stage in adulthood, and that any permanent decision to give one of them up be made with the Razian conditions satisfied.<sup>90</sup> These two stringent requirements prevent it from sliding back into a preference satisfaction account. As discussed, what individuals would prefer, full stop, and what they would prefer under the stronger conditions of having the capabilities (and, with them, the conditions of autonomy) well into adulthood, are very different things.

Thus, for the moment, I will remain agnostic on the Permissibility Claims. However, in Chapter Five it may become necessary to take the distinction seriously, and I will return to it then.

## Conclusion

This concludes my defence of the capabilities approach, taken at its broadest level as an attempt to identify the prerequisites of a flourishing human life. A solution to the PPP has been laid out, and a further objection has been met by offering two workable versions of the capabilities approach. This paper has already shown how Dworkin’s criticisms turn, at this wider level, into a version of the PPP, and are dependent, even at the narrow level, on the Paternalism and Preferences Accusations.

I am thus able to fill out the account given in the previous chapter by stating that a person has such a fundamental interest in enjoying certain broadly-defined capabilities that it is not in her ATC interest to give up one of these central

---

<sup>90</sup> It is important that the capabilities themselves, and not *only* the (arguably narrower) set of Razian conditions, are maintained up to adulthood. This is because without having had the chance at a formative stage to develop the relevant internal capacities (which may, as Nussbaum notes, require not only the enjoyment of the capability, but actually of the functioning, as a child (Nussbaum 2000b, 89-90)), the individual would often be unable to experience the capability at all. They would thus be incapable of evaluating, or appreciating it, and of making a genuine choice, *in possession of the relevant information*, to give it up.

capabilities. On the other side of the same moral coin, if a person has morally important interests, then she can be *harmed* in a morally relevant way through the undermining of those interests. I have not attempted to defend all Nussbaum's central capability categories in detail because the broadly defined categories on which my later arguments will rest are relatively uncontroversial.

I have now completed my defence of the non-intentionalist account of collectivities and morally salient potential collectivities, and argued for an expansion of the notion of moral responsibility for harm, so that sets of individuals can be collectively responsible for harm resulting predictably from the combination of their individual acts. The next task is to fill out another suggestion made in Chapter One. There, I pointed to the likelihood that fulfilment of collective duties to end harm would require *collective* action. Chapter Four will defend this claim, identifying six scenarios under which it is "better" (in the sense of better achieving the common goals, purposes or interests of the group) for a collectivity or morally salient collectivity to make certain decisions *qua* collectivity, and hand them down as restrictions on individual action, then leave pursuit of common goals or interests to the aggregation of individual decisions.

## Chapter Four

### Collectivities as Units of Agency

This chapter will ask why at least some of the actions or decisions of a collectivity should be *collective* actions or decisions, rather than the result of aggregated individual actions or decision making. The reason, I will argue, is that outcomes in such cases are better in terms of the attainment or satisfaction of the common goals or interests defining the collectivity.<sup>1</sup>

Chapter Two defended a model of collectivities as constituted by sets of individuals mutually dependent through purposes, goals or all-things-considered (ATC) interests that it only makes sense to consider as being pursued together, *whether or not they acknowledge it themselves*. There are two main reasons why decisions in such collectivities might be better made collectively. One is the inefficiency, or even counterproductivity, of leaving the outcome to the aggregation of individual acts. This will be discussed as the *Inefficiency Argument*. (Section II) The other is that individuals might not be motivated to act, *qua* individuals, in pursuit of the common goal, interest or purpose.

This could come about in either of two ways: (i) there could be a clash between what an individual sees as in his own individual interest and what is in fact collectively rational, or (ii) there could be motivation problems even where individuals are genuinely committed to some shared goal.

Both (i) and (ii) could, again, come about in various ways. Individuals might perceive what is in their own short-term interest as distinct from what is in fact in the overall interest, either because those individuals are unaware (or deny) that they are mutually dependent with others for the achievement of some interest or goal (the *Ignorance Argument*, Section III), or because there really is a clash between what is individually and what is collectively rational (*Game Theoretic Arguments*, Section IV).

In terms of (ii), individuals might fail to be motivated to act in pursuit of even a genuinely shared goal because of clashes of interest within the collectivity over how best to achieve it. (*Partial Conflict Arguments*, Section V). Even without

---

<sup>1</sup> In fact, the actions or decisions with which I am concerned are those (and only those) involved in pursuing the common interests or goals of the collectivity, as I have defined them. This restriction is one I will defend in Chapter Six.



such partial conflicts, it might not be rational for them to contribute, or be rational but in such a way that the decision-making process required to reach this conclusion places implausibly high demands on the individuals in terms of assessing probabilities. Both these possibilities will be discussed under the *Rational Altruist Arguments*. (Sections VI to IX)

I will discuss these arguments in turn. However, there are two preliminary points to stress. One is that some of the arguments have stronger implications than others. We need to distinguish between two closely connected questions: why it might be better for a group to make certain decisions collectively (and hand them down as patterns of individual action), and whether this improvement will require not only collective decision making but also the power to enforce such collective decisions on individuals. Such powers might involve sanctions (formal or informal) for non-compliance, at various levels up to, ultimately, either the threat of expulsion from the collectivity or coercive interference with an individual's property or freedom.

While some of the arguments (Game Theoretical and Ignorance Arguments) have implications for both of the first two questions distinguished above, others (Partial Conflict and Inefficiency Arguments) are less clear cut on the issue of enforceability. Rational Altruist Arguments, far from implying that enforcement is necessary, indicate rather that individuals should be motivated, without coercion, to fulfil their role in a collectively-decided process to achieve the common goal to which they are all committed.

A further question, which this chapter will not attempt to settle, is that of what further conditions would have to be met before (if ever) such coercive enforcement could be said to be *legitimate*.

The second preliminary point concerns what is meant by "better" and "worse" in this context. The next section will clarify this, and eschew some of the more worrying implications of such phraseology in any discussion of group action.

## **I. Better for whom?**

Why does this question of collective versus aggregated individual action *matter*? This chapter will argue that the outcomes of aggregated individual decision

making will, at least on occasion, be worse than the outcome of decisions made collectively (with the collectivity as the unit of agency) and handed down to individuals. But what do I mean by “worse” or “better”?

One possibility would be: better or worse for the collectivity, on the grounds of an interest assigned to that collectivity by virtue of its metaphysical status, over and above those of the individuals within it. This is *not* the claim being defended here. This project has made only very limited metaphysical claims.

In previous work, and (briefly) in Chapter Two, I have taken a position on the irreducible existence of collectivities to the extent that I have defended their expressive autonomy: the claim that they add terms to the language which enable us to express truths we could not have expressed with only the means of referring to individuals.<sup>2</sup> I have argued, further, that they meet the three criteria laid down by Graham for collectivityhood (which he takes, mistakenly, to be the same as social grouphood). That is: individuals act in ways whose significance can only be adequately captured by ineliminable reference to the collectivity itself, what the collectivity does is distinct from anything the individuals do, and its survival is relatively indifferent to the persistence of the individuals composing it at any one time.

I do not (as, for example, Sheehy does<sup>3</sup>) take a stand on the materiality or otherwise of collectivities. Moreover, I do not assign any independent interests or goals to the collectivity: any interests or goals which it can be said to have are derived directly from the goals, purposes or ATC interests of individuals. The answer to the question, better for whom? is, better for all the individuals who make up the collectivity.<sup>4</sup>

To clarify this, consider first those collectivities whose members are bound by selfish individual goals or ATC interests requiring co-operation for their achievement, then those where there is genuine espousal by individual members of some shared goal.

Regarding the former, it is helpful to borrow from W G Runciman and Sen. They use the theory of non-zero-sum non-cooperative games, in particular the

---

<sup>2</sup> Cripps 2005, 5

<sup>3</sup> Sheehy maintains that social groups are ‘compositional material particulars’. (Sheehy 2001, 121)

<sup>4</sup> Better, as will become clear, in terms of promoting the interests, purposes or goals through which they constitute a collectivity. However, I am remaining agnostic for now on whether such promotion is always in the individuals’ *all things considered* interest.

Prisoner's Dilemma, to give a substantive sense to the notion of a "general will". I will not lay out their argument, as I will be discussing the Prisoner's Dilemma in more detail later. The essential point is that, while each prisoner judges it (correctly) to be in his individual best interest to confess whatever the other does, both will be worse off if both act on this line of reasoning. As Runciman and Sen put it:

This conflict between what seems individually better and what seems to produce the best overall result contains, in our view, the essence of Rousseau's distinction between the "will of all" and the "general will". The "general will" of the prisoners, we can say, is to avoid confession, but each person's "particular will" is to confess.<sup>5</sup>

The overall result, if individuals are left to act independently, is worse in the straightforward sense of being worse for all individuals, in terms of the goals they hold as individuals, than if a decision could have been made (and stuck to) as a collectivity. (Runciman and Sen call for an 'enforceable contract'<sup>6</sup> between them.) What the general will wills is the common good.

The same holds (as Runciman and Sen add) for cases of partial conflict: that is, where all parties are better off acting as a collectivity than not so acting, but different individuals will be better off again depending on exactly *how* the collectivity acts.<sup>7</sup> The general will, Runciman and Sen argue, wills 'the just resolution of conflicting interests'<sup>8</sup> in such cases. John Smyth disputes this, claiming (although without defending his claim) that 'there is no reason to suppose that the general will does will the *just* resolution of conflicting interests. Of course, it may be said to will the resolution of them. But the *just* resolution of them is a different matter altogether.'<sup>9</sup> However, it is enough for my current point that there is a non-empty set of outcomes to be brought about by acting *qua* collectivity, each of which is more desirable to all the individuals in the collectivity than the outcome of all acting *qua* individual.

Now consider those collectivities where the common goal, purpose or interest is one that individuals genuinely espouse. Either their goal is to do some

---

<sup>5</sup> Runciman & Sen 1965, 555-6

<sup>6</sup> Ibid., 556

<sup>7</sup> I will return to such cases with the discussion of the Battle of the Sexes in Section V.

<sup>8</sup> Runciman & Sen., 558

<sup>9</sup> Smyth 1972, 429. See Runciman & Sen 1974 for their response

particular thing *with that particular group* and/or the common goal or interest is one an individual cannot have alone – it only makes sense as part of a collectivity. Nonetheless, the moral force of the goal or interest derives from its contribution to *individual* lives. In such cases, the collectivity is constituted by *individuals* who have as a goal (or in whose ATC interest it is) to be *part* of a set of individuals with this shared aim.

For example, the goal might be that of being one of a set of individuals constituting an orchestra: a set which, *qua* orchestra, can play beautiful music. This may be simply because the individual wants to do something which cannot be done except as part of a group (she wants to take part in concerts performing classical works, and so to be part of an orchestra – any orchestra) or because she has the goal of doing something with certain particular individuals. (Suppose four musician friends choose to form a string quartet: each, we can assume, has the goal of producing the beautiful music with those other people, but it is still a goal the *individual* has.)

I hope this has clarified that when I speak of an outcome being better for a collectivity, I mean better *in terms of promoting the common goals or interests of the collectivity, which goals or interests derive from the goals, purposes or ATC interests of the individuals within it*. For the rest of this chapter, I will defend the claim that outcomes will be better in this sense if at least some collective results are the outcome of decision making *qua* collectivity, rather than the aggregation of decisions made *qua* individual.

## **II. Inefficiency arguments and the Discursive Dilemma**

This section will examine what might be called “pure efficiency” arguments for collective decision making, by making a general point about efficiency, then appealing to a particular case: the Discursive Dilemma. Here, we are setting aside any potential problems regarding individual commitment to, or acknowledgement of, shared goals or interests, and highlighting simply the potential for loss, in terms of achievement of shared goals or satisfaction of common interests, if the individuals all act independently.

There are infinitely many scenarios, in collectivities big or small, where it is necessary to co-operate in deciding what to do, because decisions made

individually and aggregated could undermine the progression of the collectivity towards its common goals, interests or purposes. On a trivial level, suppose you and I want to spend the evening together. (This makes us, at least in the short term, a collectivity.) Neither of us has any strong feelings about what exactly we do, so long as we are together. However, we still need to come to some agreement: if we don't coordinate properly, and I turn up to the cinema while you go to the restaurant, then our shared aim of doing something together is foiled.

Scenarios such as these explain the establishment of many conventions. Examples are: driving on the left hand side of the road, walking up or down the left hand side of an escalator, using three tugs on the rope in rock climbing to indicate it is safe for one's partner to start climbing, flashing lights to warn oncoming traffic of some obstacle on the road ahead of them. It does not matter to the individuals exactly which side of the road or escalator, what number of tugs, etc. is chosen, but each collectivity has to establish some convention or the outcome will be sub-optimal for all.

Equally, consider Derek Parfit's example of an act-utilitarian society: if we are all '*pure do-gooders*', motivated only by a straightforward act-utilitarianism, we might not actually maximise happiness.<sup>10</sup> Much of our happiness, he notes:

comes from having, and acting upon, certain strong desires. These include the desires that are involved in loving certain other people, the desire to work well, and many of the strong desires on which we act when we are not working. To become pure do-gooders, we would have to act against or even to suppress most of these desires. It is likely that this would enormously reduce the sum of happiness.<sup>11</sup>

Again, the overall outcome would be sub-optimal unless the collectivity decided together how best to achieve the goal, and acted accordingly (as rule utilitarians).

Following these general points, a more specific efficiency argument can be offered. This derives from a problem in jurisprudence, the doctrinal paradox, and has been adapted and generalised by Pettit as the discursive dilemma.

The discursive dilemma arises because different action may be taken if a collectivity decides on the basis of individual voting on premises (and the logical

---

<sup>10</sup> Parfit 1984, 27

<sup>11</sup> Ibid., 27

conclusions to be drawn from these) than if it had decided on the basis of individual voting on conclusions. Consider the following scenario. A group of environmental activists (a voluntary, intentional collectivity) are concerned with the threat a proposed new building would make to the survival of a local flock of rare birds. They are considering whether to campaign actively against it by mounting a protest in the Town Square.

Adapting Pettit's 'conjunction case',<sup>12</sup> suppose the group is composed of three members, A, B and C, and they are deciding whether to campaign on the basis of two issues:

- (i) whether there is a serious risk to the survival of the bird flock from the new building;
- (ii) whether the campaign has a high chance of preventing it.

The idea is that, if and only if someone answers yes to both (i) and (ii), she votes in favour of the campaign. Suppose that the matrix of answers is as shown below.

|   | Serious risk? | High chance of prevention? | Campaign? |
|---|---------------|----------------------------|-----------|
| A | Y             | N                          | N         |
| B | N             | Y                          | N         |
| C | Y             | Y                          | Y         |

This is not a unique case. It can arise, Pettit notes, whether the following hold:

- a. there is a conclusion to be decided among a group of people by reference to a conjunction of independent or separable premises - the conclusion will be endorsed if relevant premises are endorsed, and otherwise it will be rejected;
- b. each member of the group forms a judgement on each of the premises and a corresponding judgement on the conclusion;
- c. each of the premises is supported by a majority of members but those majorities do not coincide with one another;
- d. the intersection of those majorities will support the conclusion, and the others reject it, in view of a; and
- e. the intersection of the majorities is only a minority in the group as a whole.<sup>13</sup>

<sup>12</sup> Pettit 2001, 272

<sup>13</sup> Ibid., 274. There is, moreover, a parallel 'disjunction case' which I will not discuss here. (274-6)

The dilemma arises because, if a vote is taken on individual conclusions, the campaign will not go ahead. If, however, the members of the group vote on the premises, then as a group, they answer yes to both (i) and (ii) and so, logically, should support the campaign.

On the face of it, neither appeals. As Pettit sums up, 'going the conclusion-driven way means adopting a course that is inconsistent with the premises endorsed by the group and going the premise-driven way means adopting a course that a majority individually reject'.<sup>14</sup> Pettit argues that the premise-driven approach is superior on the grounds that it enables individuals to question group or collective (often government) decisions by demanding to know the reasons behind them. As far as this chapter is concerned, the case can be made directly relevant by considering it as follows.

Suppose, rather than voting on whether to campaign, the individuals are simply turning up to a protest in the Town Square (or not) as they judge appropriate. Suppose, moreover, that the group has 90 members, each of whom knows that the local authorities will not even register that there has been a protest (and so consider whether to act on its demands) unless there are at least 60 protesters. Of the 90 environmentalists, 30 reason as A in the matrix, 30 as B and 30 as C. The result of this, effectively a conclusion-led decision, is that there are not enough protesters even to get noticed, even though 60 members of the group believe that a protest would be effective at preventing the building, and 60 believe that the building would seriously threaten the birds.

I suggest that this *must* be a sub-optimal outcome, in the sense of failing to optimise pursuit of the group's goals, *if* we make the assumption:

K: the majority verdict on each premise will be the right one (or be most likely to be right)

If K holds, the best way to achieve the common goal of the group (for now assumed to be the relatively narrow one of protecting local birds) is to assess individual views on the premises and draw the logical conclusion, collectively,

---

<sup>14</sup> Ibid., 274

rather than leave it to individuals to draw their own conclusions and vote with their feet.

K can be considered in relation to two general classes of premise. The first is factual premises. This includes both premises in the example above (although they may, by their nature, have to be estimates). In this case, various circumstances improve the chances of K holding. One is that the individuals are influenced by scientific work by unbiased experts (rather than, say, by tabloid headlines or what the local preacher says). If individuals have familiarised themselves with scientific work on the subject, accepted where there is consensus and sifted through the evidence to make an educated decision where there is not, then there are grounds for optimism that, so far as modern science will get it right, then a majority vote will reflect this.

Secondly, there are cases where a judgement call on at least one of the premises must be value-based (or at least relative to individuals). Suppose that there is no doubt of the risk to the birds, but that (i) is replaced with:

(i') whether it is more important that the birds be preserved or that resources be devoted to replanting certain rare types of tree.

Then it seems that there is no other way of deciding the issue than debate followed by (if no consensus can be reached) a vote. Of course, if divergence is sufficient and on sufficiently "key" issues, then with many voluntary groups the solution will be for those favouring the priority of path X to form a separate collectivity to those favouring path Y. However, in some cases, this is not an option (or is too drastic a response). Consider for example the group of travellers committed to each other through sharing a car, and having to choose between two possible sights, only one of which they have time to visit. Ideally, it seems, a consensus would be reached, but a decision has to be made, and if debate fails, then it seems that a vote is the closest the collectivity can get to a "group" view.<sup>15</sup>

Opponents of the claim above might try to identify cases where K holds but nonetheless the conclusion-driven approach yields a better outcome. However, intuitively and drawing on remarks by Pettit, such scenarios are likely to take two possible forms: the two premises are somehow inseparable (or one is so

---

<sup>15</sup> That is not to say, of course, that there would not be other available ways of reaching a verdict, for example appointing the car driver as a dictator.



important as to outweigh the other), or the aim of the decision process has been incorrectly clarified.<sup>16</sup>

Suppose, for example, either that: unless there is a reasonable chance of success there is no point carrying out the protest whatever the seriousness of the risk, *or* that the more serious the risk to the birds, the higher the chance of success. In either case, each individual's considered judgement will be on the conclusion rather than independently on the premises. Alternatively, suppose that what matters to the group is not reaching the outcome most likely to be accurate. Rather, what is required is that the campaign *only* take place if there is no room for doubt on either condition. In that case, a unanimous "yes" on each premise would be required.

In none of these scenarios is there a genuine discursive dilemma: either the premises are not independent, or the process of reaching a conclusion is not as laid out by Pettit.

Moreover, even where K does not hold, and so a premise-driven approach will not be optimal, this does *not* mean that a conclusion-driven one will be any better. There is no reason to assume that individual mistakes will cancel each other out *unless* the premises are in some way interdependent – in which case there is not a genuine discursive dilemma.

Consider the activists, with respect to premises (i') and (ii). Given that (i') is value-based, we can assume the unreliability of the majority to be with regards to (ii). The scenario might play out in various ways, but the following exemplify the main possibilities:

1. Members of faction (A) are all experts on the statistics regarding previous environmental demonstrations and their success rate. Members of (B) and (C) have no such knowledge.
2. None of the members of the collectivity have any information on the success rate of demonstrations, but there is an independent body with considerable expertise, to which appeal could be made.
3. None of the members of the collectivity have any information on the success rate of demonstrations.

---

<sup>16</sup> In relation to his example of (not) granting a retrial. (Pettit 2001, 286-7)

In none of these cases is appeal to the conclusion of each individual any more likely to yield the optimal outcome in terms of the goals of the collectivity than appeal to the premises. Rather, while retaining appeal to the majority on premise (i'), it would be better in Scenario (1) for the collectivity to defer to the expertise of faction A on premise (ii), or in Scenario (2) to appeal to the independent body for advice. In Scenario (3), there seems to be no point in appealing to (ii) at all, as a factual premise. It could be rewritten as a more evaluative one – is there any point taking a chance on the unknown probability of being successful – in which case K might well hold and a premise-driven approach would, again, be most likely to be optimal.

I have discussed this at more length elsewhere, but need no more detail here.<sup>17</sup> The point to be extracted in support of the Inefficiency Argument is that a situation has been identified in which the aggregation of individual decisions or individual actions can fail to achieve an optimum outcome in terms of the goals of the collectivity – and that there is no reason to suppose, in those cases where an aggregation of premises would not be the optimum way to make a collective decision, that an aggregation of conclusions would do any better. Rather, some other way of making a collective decision is called for.

There are two points to note before leaving this section. Firstly, whether enforceability is likely to be required will depend on how well individuals understand and accept the kind of reasoning provided in this section. For example, do they accept the need, ultimately, for a vote on value-based premises where a decision has to be made?<sup>18</sup> Do they accept that their own factual views may be mistaken? Secondly, many of the cases discussed in later sections (such as the Common Land Farmers or the Rational Altruists) can be supplemented with efficiency-based arguments.

---

<sup>17</sup> Cripps (2006 conference paper): “The Discursive Dilemma and Collective Decision Making”

<sup>18</sup> I am not advocating the imposition of ideological views at a state level. Chapter Six will make this clear. Much of this applies at the level of the voluntary collectivity where an individual could, ultimately, opt out. Where this does apply in non-voluntary collectivity, it is likely to be because some decision *has* to be made, for example on whether to preserve some piece of land as wilderness or build houses on it. (In at least some such cases, although this goes beyond the scope of this project, I would suggest that the interested parties include non-humans, who ought to be represented in the decision process, even though they cannot express an opinion themselves.)

### III. Ignorance arguments

The next argument for collective decision making focuses on those individuals who do not or cannot acknowledge their membership of a collectivity.

The account defended in Chapter Two expressly allowed for such cases. It required that the individuals be mutually dependent through their purposes, goals or ATC interests – *whether or not they acknowledge it themselves*. I spelled out, moreover, that this includes both goals/purposes/interests which, by their nature, can only be pursued together, and goals/purposes/interests which in the circumstances can only be pursued together.

I made clear that I wanted to include both of the following types of cases: a set of individuals whose goal is to do something *together* (for example, the friends climbing Monte Rosa), and a set of individuals with individual goals who do not care whether everyone else succeeds in achieving their goal, when the circumstances are such that they must all co-operate in order for each to achieve his own goal (as in the *Lost* example). This leaves open the possibility that individuals might either refuse to acknowledge an ATC interest; or acknowledge an interest, purpose or goal but deny that it required collective action.

With regard to the former, Chapter Three defended an understanding of human flourishing according to which we all have certain fundamental interests (central capabilities). For example, all members of a nation state are assumed to have such an interest in living in a safe, peaceful society, even if terrorists might reject those interests.<sup>19</sup> Similar points will be made in the next chapter, regarding the curtailment of climate change. The case can be made even less controversially by consideration of those (such as the mentally handicapped, or infants) who cannot be aware of their own interests.

The latter possibility (acknowledgement of an interest but not of its being a common interest) could arise in the *Lost* scenario. Each of those washed up on the island might erroneously be convinced that he had the wherewithal to catch food, build shelter, etc. for himself, by himself.

In both scenarios, at least some individuals within the collectivity cannot be relied on to act in pursuit of the common interest, purpose or goal. In such cases,

---

<sup>19</sup> See Chapter Two, Sections IV and VII

collective decisions, restricting the behaviour of all (through rules or less formal conventions) are necessary for the furthering of that end. Clearly, the need for enforceability goes alongside the need for such collective decision making: without acknowledgement by individuals that they are even part of a collectivity, there is hardly likely to be voluntary compliance with collective decisions.

A sub-case of this is those individuals who know, at some level, that they have a shared goal, purpose or ATC interest with certain others, but bury their heads in the sand. Consider the following example. A village, made up of various farmers, has been built on land reclaimed from marshland. Considerable maintenance of the land is required to prevent long-term risk of flooding. Carelessness by any farmer in maintaining and protecting his plot is a serious risk to the long-term future of the whole village. The farmers are rendered an involuntary collectivity by virtue of their mutual dependence through the common interest: saving the land. However, the costs of such maintenance in the short term are significant.

Consider the following decision matrix faced by each farmer.

*Head-in-the-Sand Farmers*

|  | This farmer carries out maintenance   | This farmer does not carry out maintenance   |
|--|---|--|
| Other farmers carry out maintenance        | A: Low profits for all farmers this season, but the land is saved.  | B: Very low profits for other farmers this season, extremely high profits for this farmer this season. Land almost certainly unusable in the long term |
| Other farmers do not carry out maintenance | C: Very high profits for other farmers this season, extremely low profits for this farmer this season. Land almost certainly unusable in the long term. | D: High profits for all this season. Land almost certainly unusable in the long run.   |

B is a worse outcome than A, because the land cannot be saved in the long term unless *all* farmers carry out maintenance. However, each farmer might well ignore this, *especially* (given that D is much better for him than C) if he has no guarantee that others will co-operate.

The farmers might understand but ignore the problem of the decline of the land in order to boost immediate profits, hoping vaguely, since it is not impressed upon them on a daily basis, that the problem will go away. Or they might go further, and adopt unfounded factual beliefs. As David Papineau has argued, while individuals faced with potential conflict in their overall set of desires and beliefs will need to adjust one or the other (assuming the evidence of that clash to be forced regularly upon them), they could do this in two ways: by adjusting their desires, *or* by adopting beliefs that match their existing desires.<sup>20</sup> In this case, if they *are* constantly reminded of the situation but really don't want to make the effort to do the maintenance, they might, like smokers denying the link between smoking and lung cancer, deny the scientific evidence regarding the marshland.

It could be argued that the best solution in such cases is education and repeated spelling out of the situation, so it is harder to ignore or deny. Clearly, this is an advantage whether or not decisions are made collectively. Whether they are expected to follow rules or conventions, or act voluntarily, as individuals, to bring about the best outcome, it is far more likely that a solution will be unproblematic and stable if the individuals understand why they are acting in this way. But it seems implausible that education alone would always be enough. Humans, in general, are remarkably adept at burying their heads in the sand until it is too late.<sup>21</sup> General attitudes to environmental risks are a case in point.

#### **IV. Game theory arguments**

Such cases bring us to the second case for collective decision making. The non-intentionalist model of a collectivity allows for "common goals" which are individual goals requiring the co-operation of the whole of the set for their

---

<sup>20</sup> Papineau 1978, 168

<sup>21</sup> I am putting aside the question of who is to direct collective action if *all* members of the collectivity are genuinely ignorant of the situation. The point in the environmental case is that some of us are now aware.

achievement. However, this opens up the possibility of collectivities whose members, *even if they acknowledge the situation*, will be rationally motivated, *qua* individuals, to act in a way which yields an overall lower-than-optimum (or worst) outcome.

This can be illustrated with two Game Theory dilemmas: the Prisoner's Dilemma and Chicken. These, while identifying small, involuntary collectivities, are most useful in casting light on wider dilemmas which call for collective decision making.

Let us start with the Prisoner's Dilemma. Two prisoners, call them Jack and Jill, are to be questioned (separately) about a crime. If neither confess, then they will both be given two years in prison (second best outcome, S, for both). If Jill confesses and Jack doesn't, she will be let off altogether (best outcome, B) but he will go down for ten years (worst outcome, W). The same applies in reverse. If both confess, they will both spend five years in prison (third best outcome, T, for both). The decision matrix faced by Jill and Jack is:

#### *Prisoner's Dilemma*

|      |               | Jack          | Jack    |
|------|---------------|---------------|---------|
|      |               | Don't confess | Confess |
| Jill | Don't confess | S,S           | W,B     |
| Jill | Confess       | B,W           | T,T     |

Whatever Jack does, Jill will be better off confessing (B is better than S and T is better than W). Confessing is a dominant strategy.<sup>22</sup> This is a symmetric game, so the same applies to Jack. Thus, both will be rationally motivated to confess, and the resulting outcome will be (T,T). The paradox, as Henry Hamburger spells out, is:

- 1) Each player has a dominant strategy
- 2) If both players use the dominant strategy, the result is worse for *both* players than if both instead had used their *dominated* strategy.<sup>23</sup>

<sup>22</sup> I am borrowing my terminology from Hamburger 1979.

<sup>23</sup> Hamburger 1979, 78

Jill and Jack constitute a small, involuntary collectivity on my account. They have individual goals (minimising time in prison) which, under the circumstances, require co-operation to be achieved: by agreeing not to confess and sticking to that, they would achieve outcome (S,S) which is better for both than (T,T). The problem is that, so long as they are individuals acting *qua* individuals, and the pay-offs are as they are, they cannot trust each other. Even if they agree not to confess, each will be motivated to renege, and so boost his own pay-off. (In Game Theory terms, the cell (S,S) is not an equilibrium because either player could, acting independently, improve his or her outcome.<sup>24</sup>)

The goal can be achieved by making the collectivity the unit of agency, with the individuals in some way bound by the joint decision. Again, the need for enforceability is bound up with the need for collective decision making. In different collectivities, this can take different forms: conventions associated with social disapproval for those who break them, or rules and penalties. Jack and Jill might ensure that the rest of their friends and associates know of the deal, so that anyone who breaks it will be socially excluded. Or they might have buried their stolen loot in a safe with a combination half of which is known to each, so anyone who breaks the "rule" and confesses will be penalised. The point is that the decision has to be made collectively, with individuals bound to that collective verdict, for the common goal (in this case the individual goals) to be achieved.

This artificial case has far-reaching implications, as can be seen using a related example: the Common Land Farmers. A number of farmers all want to graze their cows on some village land, but if they all keep on adding cows, the land will quickly become unusable. They are mutually dependent through each of their long term interest in retaining use of the land, rendering them a collectivity. However, there is a rational incentive for each farmer to free ride. Consider the following decision matrix.

---

<sup>24</sup> Ibid., 85

*Common Land Farmers*

|  |  |  |
|--|--|--|
|  | This farmer reduces cow numbers  | This farmer does not reduce cow numbers  |
| Other farmers reduce cow numbers       | A: Somewhat low profits for all farmers this season, but the land is usable indefinitely.  | B: Very low profits this season for other farmers, extremely high profits for this farmer this season; land usable indefinitely. |
| Other farmers don't reduce cow numbers | C: Very high profits for other farmers this season, extremely low profits for this farmer this season; land unusable beyond this season. | D: Somewhat high profits for all farmers this season; but the land is unusable beyond this season.                               |

This is not a pure Prisoner's Dilemma: the case is not symmetric seen as one farmer versus many. However, the situation is the same in that this farmer (and so any individual farmer) has a dominant strategy: don't reduce cow numbers. B is a better outcome for him and A, and D better than C. But if all farmers reason this way and don't reduce cow numbers, the outcome will be D, which is worse for all than A. Collective, coordinated (enforceable) action is required for the achievement of common ends.<sup>25</sup>

This is a very real problem when considered in the light of environmental issues, which require some small but significant action more or less across the board. (The next chapter will make the case for at least most of humanity to constitute a collectivity on the grounds of a common interest in environmental protection.) It is one, I would argue, which calls for collective action and collective decision making. Consideration of the situation as analogous to a game theoretic problem makes this clear.<sup>26</sup>

Parfit suggests two ways out of such situations: political solutions and psychological (including moral) solutions. Perhaps, as he outlines, individuals

<sup>25</sup> Note that even if farmers did all act independently to reduce cow numbers, this might be inefficient compared with (say) a rota system.

<sup>26</sup> The use of game theory to explain the problems posed by environmental degradation is nothing new. See Hardin 1968 and Gardiner 2002



could be made: trustworthy (so that they could all promise not to free ride on the condition that the others do the same), reluctant to be free riders, Kantians (so that each does only what he could rationally will all to do, so none free-ride), or sufficiently altruistic not to free-ride. Alternatively, the situation could be changed so that the free-riding option was no longer available, or other options relatively more attractive.<sup>27</sup>

Past experience suggests the implausibility (especially with large groups) in relying on all individuals becoming morally motivated in one of the four ways. This leaves us with the political solutions. For Parfit, these are roughly categorised as solutions which change the situation rather than the individual – either by making certain options inaccessible or by adjusting the payoffs so that the options which lead to the best overall outcome become the most attractive ones to individuals.<sup>28</sup>

This translates into a need for a collective decision to be handed down to individuals as enforceable rules, or conventions, more or less formally, depending on the collectivity. However, as Parfit also notes, it will be much easier (sometimes only be possible) to achieve stable political solutions if individuals have at least some awareness of and are at least to some extent moved by the moral motivations.

The second case to consider is “Chicken”. Two young men are driving towards a cliff edge. Whoever brakes first will “lose” the game, but it would be a much worse outcome for neither to brake. The matrix is:

*Chicken*

|       |                   | Boy 2       | Boy 2             |
|-------|-------------------|-------------|-------------------|
|       |                   | Brake first | Don't brake first |
| Boy 1 | Brake first       | S,S         | T,B               |
| Boy 1 | Don't brake first | B,T         | W,W               |

Unlike with the Prisoner's Dilemma, there is no dominant strategy for either player. If Boy 2 brakes first, Boy 1 will be better off not braking until just

<sup>27</sup> Parfit 1984, 62-6

<sup>28</sup> Ibid., 63-4

afterwards; if Boy 2 doesn't, then for Boy 1 not to would mean the worst outcome. On a maximin strategy (that is, make the worst possible outcome as good as possible), either boy would brake, meaning that the worst possible option would be that he braked and the other didn't. But each, assuming the other would reason this way, would have reason to make his own situation better off by not braking (the cell (S,S) is not an equilibrium). Both reasoning thus, however, would mean the worst outcome for both (W,W).<sup>29</sup>

On the non-intentionalist model, the boys form a (very unwilling, ad-hoc) collectivity. The only way, in practice, to avoid disaster, is for both to brake. However, this can only be achieved by each committing himself to act as a member of the collectivity, abiding by a collective decision. Of course, in the original case this would be entirely contrary to the spirit of the game and would never happen.<sup>30</sup> But it is echoed in global politics.

Hamburger, writing in 1979, develops a model of what he calls "Nuclear Chicken".<sup>31</sup> There are two superpowers, US and SU. Each has the choice of accommodation or aggression with nuclear weapons.

Assume that if both sides choose accommodation, then the status quo will continue... If one side is aggressive... and the other accommodates, then the accommodating side would suffer strategic and perhaps trade and diplomatic losses (-2), while corresponding gains would accrue to the aggressor, though in lesser amount (+1) because the resulting increase in world tension might generally interfere with trade and diplomacy. If both sides choose confrontation (with no willingness to accommodate), then the result is taken to be a nuclear exchange, presumably by far the worst outcome for each (-10).<sup>32</sup>

The matrix is thus:

|    |               | US            | US         |
|----|---------------|---------------|------------|
|    |               | Accommodating | Aggressive |
| SU | Accommodating | 0,0           | -2,1       |
| SU | Aggressive    | 1,-2          | -10,-10    |

<sup>29</sup> For a full discussion of this case, see Hamburger 1979, 83-7

<sup>30</sup> In 1955 film *Rebel Without A Cause*

<sup>31</sup> Hamburger 1979, 20-21

<sup>32</sup> Ibid., 21

Or:

|    |               | US            | US         |
|----|---------------|---------------|------------|
|    |               | Accommodating | Aggressive |
| SU | Accommodating | S,S           | T,B        |
| SU | Aggressive    | B,T           | W,W        |

Today, with various countries likely to have nuclear weapons, the model would be a wider one, but the point remains the same. No player has an incentive, *qua* rational individual and assuming the others to be rational, to be accommodating. But if all are aggressive, the result is nuclear war, and total disaster, not just for the respective players, but for the whole world.

As will be discussed in the next chapter, I would defend the claim that at least most of humanity (including all of my generation and below) is a collectivity. The governments of the nuclear-weapon-owning countries would thus constitute an active sub-collectivity. As with the Prisoner's Dilemma cases, the rational thing to do, as a collectivity, is to make a collective decision to which all players (all members of the collectivity) are bound.

## V. Partial conflict arguments

The last two sections focused on clashes between what the individual perceives as in her own interest, and what is in fact in the collective interest (and so, ultimately, in hers). From now on, the assumption will be that individuals *are* genuinely committed to some shared goal or purpose. However, motivation can still arise.

Partial conflict may arise within the collectivity, which can only be settled by a collective decision. Consider the Battle of the Sexes: a couple want to spend the evening together. However, she wants to stay in; he wants to go out. Each would rather do what the other wants, than separate. The decision-matrix is:

*Battle of the Sexes*

|     |         |         |        |
|-----|---------|---------|--------|
|     |         | Her     | Her    |
|     |         | Stay in | Go out |
| Him | Stay in | S,B     | W,W    |
| Him | Go out  | T,T     | B,S    |

One will have to make a sacrifice (do what the other wants), or the outcome will be worse for both. The matrix can be redrawn in these terms, as another Game Theory case, Hero.

*Hero (Battle of the Sexes)*

|     |                 |                 |           |
|-----|-----------------|-----------------|-----------|
|     |                 | Her             | Her       |
|     |                 | Don't sacrifice | Sacrifice |
| Him | Don't sacrifice | T,T             | B,S       |
| Him | Sacrifice       | S,B             | W,W       |

Coordination is essential, as Hamburger notes, because it is essential that one yield but both do not.<sup>33</sup> On occasion, deciding separately, they might happen on (S,B) or (B,S) outcomes, but for this to happen reliably requires the decision to be made as a pair, for the pair. Moreover, each player, knowing that the other will be better off yielding, if he or she does not, will be motivated not to yield. This, however, would mean a second-worst outcome of each doing as he or she chose, but separately.

Again, there is a motivation to free ride. This becomes clearer in wider cases. Suppose several children are playing catch, and the ball lands in the neighbour's garden. If no child climbs over the fence to get it, the game will stop (second-worse, or third best outcome for all). If one child gets it, the game can go on (second-best outcome for him, best for the rest). But if all rush to the fence and climb over, they will make such a noise that the neighbour will notice and confiscate the ball (worst outcome for all). Each child has a motivation to free ride and hope another will volunteer, but if all think this way, the outcome is second worst for all. Even if they are not free riders, they need to decide as a

---

<sup>33</sup> Hamburger 1979, 89

group that only one person (and which) will climb the fence, because for all to do it would be counterproductive.<sup>34</sup>

Again, it is better if the collectivity, rather than its individual members, is the unit of agency. Note, however, that we cannot draw the same clear cut implications in terms of enforceability as in the previous two sections. If a collective decision has been made, the individual whose “turn” it is to sacrifice (or who, through some method such as a lottery, has been picked as the sacrifice) faces, on the surface of it, a choice between complying, and getting her own second best outcome, and not complying, which, given that others will not be sacrificing, results in the worst outcome. Thus, it seems to be irrational to cheat (although that is no guarantee that certain individuals will not do so, on the chance that someone else will sacrifice themselves at the last minute rather than see the collective project go under).

## **VI. Rational Altruist Argument A: Parfit’s classic case**

The next four sections will be devoted to the two Rational Altruist Arguments. In one scenario (Sections VI and VII), while it is rational for individual altruists to contribute, it is far from certain that they will do the calculations needed to work this out. In other words, individuals, *qua* individuals, may well not be motivated by what is rational. In the other scenario, it is by no means certain that it *is* rational for individuals to contribute.

Assume all members of the collectivity are genuinely committed to its common goals. The basic Rational Altruist Dilemma, which is most often cited as an objection to act-utilitarianism, goes as follows. Even if individuals are committed to the common good, circumstances arise where the collectivity is sufficiently large that the difference made by an individual, *qua* individual, is either none, or so small as to be imperceptible. Why, then, should any given individual, however committed she is to the goal, contribute to its pursuit? Surely she could decide rationally that, because the difference she makes is negligible (or none at all), her efforts would be better spent elsewhere. But, of course,

---

<sup>34</sup> A similar argument can be made in situations where the top-left and bottom-right outcomes are switched, so that the result of no-one sacrificing himself is worse than all doing so, for example Hamburger’s case of ‘several people standing on a dock as someone is about to drown’ (1979, 89)

everyone could reason the same way, resulting in a sub-optimal outcome (failure to reach the goal). Many environmental efforts are vulnerable to this kind of motivational difficulty.

Parfit and Michael Otsuka discuss the Rational Altruist Dilemma using the example of a large number of wounded men lying in the desert. Call this number a thousand, but it could be made arbitrarily big until the example convinced. There are an equal number of altruists. Each has a pint of water. If all these pints are put into a water cart and divided between the men, each will have his suffering significantly reduced. (We assume that there is no benefit besides this alleviation.) However, the addition of each pint to the cart, increasing each man's share by a thousandth of a pint, does not reduce his suffering perceptibly. Why should each individual rational altruist add his pint?

There is an important difference here between cases of very small benefits, and those where benefits genuinely are imperceptible. If the benefit of each altruist's extra pint in the cart to each wounded man was tiny, but just noticeable, he could appeal, on consequentialist reasoning, to the fact that, while his benefit to each man was small, added up across the thousand men, it is much larger, and justifies his effort.

Assuming, however, that there is no perceptible reduction of suffering to each man from the extra thousandth of a pint, to justify the choice of the rational altruist to contribute *without* appealing to the collectivity as the unit of agency, one of two paths must be taken. It is necessary either to argue that there *is* a benefit, even if it is an imperceptible one, or to find a way of justifying participation without it. Parfit attempts the latter, although he does believe there can be imperceptible benefits in the sense that pain can be reduced by imperceptible amounts.<sup>35</sup>

Otsuka attacks his argument, defends the possibility of imperceptible benefits and so finds, he thinks, a rational motivation for the altruist. I agree (this section) with Otsuka's rejection of Parfit, but draw a distinction in the next section between benefits and what I will call benefit ingredients. This means that while, ultimately, the rational agent should be motivated to contribute, this is not for the reasons Otsuka offers. It requires, moreover, a degree of rational calculation of

---

<sup>35</sup> 'I believe that someone's pain can become less painful, or less bad, by an amount too small to be noticed.' (Parfit 1984, 79)

expected utilities by individuals that is probably unrealistic and certainly uncomfortably unreliable.

Parfit argues that the rational altruist can appeal to a variant on consequentialism:

(C10) When (1) the best outcome would be the one in which people are benefited most, and (2) each of the members of some group could act in a certain way, and (3) they would benefit these other people if *enough* of them acted in this way, and (4) they would benefit these people *most* if they *all* acted in this way, and (5) each of them both knows these facts and believes that enough of them will act in this way, then (6) each of them ought to act in this way.<sup>36</sup>

(Note that while this appeals to what everyone in the collectivity can do together, it retains the individual as the unit of agency.)

The problem, Otsuka argues, is that this has two unsatisfactory implications. Firstly, he invites us to suppose that, as well as the thousand men, there are another thousand, less wounded and less thirsty (although still thirsty). Each altruist, as he decides whether to add his pint to the water cart for the benefit of the thousand very thirsty men, knows he also has the option of giving his pint to one of the less thirsty ones, and giving him the significant benefit of a whole pint of water. If all do this, then the overall benefit will be less (we assume that there will be greater benefit from reducing the suffering of the worst-off men).

Suppose we appeal to Parfit's C10. It turns out that it is not possible to specify a set of actions that will *most* benefit people. The outcome in which all give to a slightly thirsty person is worse than the one where all give to the water cart but, because the benefit to the very thirsty of each extra pint in the cart is imperceptible, it can be said of any outcome involving some contribution to the water cart, that it would be better if one person took their contribution away and gave it to a moderately thirsty person. Thus, Otsuka argues, the Parfitian has to commit himself 'to the belief either (1) that each outcome is less beneficial than itself, or (2) that "less beneficial than" is not transitive'.<sup>37</sup> This is Otsuka's Paradox of Group Beneficence.

---

<sup>36</sup> Parfit 1984, 77

<sup>37</sup> Otsuka 1991, 141

Parfit might (indeed, he later does), reject C10 and appeal instead to C7:<sup>38</sup>

(C7) Even if an act harms no-one, this act may be wrong because it is one of a *set* of acts that *together* harm other people. Similarly, even if some act benefits no-one, it can be what someone ought to do, because it is one of a set of acts that together benefit other people.<sup>39</sup>

This might allow him to avoid this objection, but it fails to avoid Otsuka's second implication. Consider the case of overdetermination. Suppose, that is, that there are one thousand altruists already and the water cart is full. On C10 or C7, an additional altruist would still have a motivation to add his pint, to become part of the group benefiting the wounded. To avoid this, Parfit adds:

(C13) Suppose that there is some group who, by acting in a certain way, will together benefit other people. If someone believes that this group either is, or would be if he joined, *too large*, he has no moral reason to join this group. A group is *too large* if it is true that, if one or more of its members had not acted, this would not have reduced the benefit that this group gives to other people.<sup>40</sup>

But, as Otsuka points out, this does not allow the rational altruist to reach a conclusion unless we assume that there can be imperceptible benefits. If they can't, then if the group with 1001 altruists is too large, then so must be the group with 1000, because the benefits it gives are no smaller, and so on.

## VII. Rational Altruist Argument A: imperceptible benefits

Let us return, then, to the possibility that there are such things as imperceptible benefits and harms, which can be "counted" in calculation of options by rational agents. They could, therefore, be summed in the same way as very small benefits/harms (although this is not the line that Otsuka takes; rather he rejects

---

<sup>38</sup> Ibid., 143 (He cites Parfit: "Comments", *Ethics* 96, 1986)

<sup>39</sup> Parfit 1984, 70

<sup>40</sup> Ibid., 83



group beneficence in favour of his restricted principle).<sup>41</sup> Otsuka affirms what he calls the Whole Is No Greater proposition:

W: If  $n$  people contribute toward the realisation of some benefit, and nothing other than the contributions of these people affects the amount of benefits secured, then the difference in benefits, if  $n-1$  rather than  $n$  contributed, added to the difference if  $n-2$  rather than  $n-1$  contribute, added to the difference if  $n-3$  rather than  $n-2$  contribute, and so forth until we reach  $n-n$ , must equal the quantity of the benefit secured if  $n$  people contribute.<sup>42</sup>

This, he argues, can be formalised as:

$$F: (B_n - B_{(n-1)}) + (B_{(n-2)} - B_{(n-3)}) + \dots + (B_1 - B_0) = B_n$$

He points to Regan's argument that this is a mathematical truism, since all but  $B_n$  and  $B_0$  are cancelled out, and  $B_0$  (the benefit of no people contributing) is, by hypothesis, zero.<sup>43</sup> I am not going to argue with this obvious maths, but I do not think that it can be used in the way Otsuka wants. By sketching an argument for the notion that imperceptible benefits *must* be benefits, it is possible to bring out the confusion behind the debate.

Such an argument might go as follows: "The claim is that by giving an extra drop of water to a very thirsty man, one is not benefiting him. However, there seems to be a clear sense in which this *is* benefiting him: the same sense as that in which adding a gram of flour to a cake is increasing the amount of flour, albeit by an amount so small as to be imperceptible in terms of the overall cake.

"We are not dealing with pre-set units: units are imposed on the scale of increasing benefit (most importantly, perhaps, the pivotal point at which a man is *aware* of being less thirsty) so that we can measure it, in exactly the same way that measurements of weight are divided into tonnes and kilograms, and grams, and if necessary into decimal divisions of grams. Our weighing scales may give us readings no closer together than 5g. If we add 1g, it will not show up. But that

---

<sup>41</sup> "[I]t is morally best for an individual to maximise whenever such maximising neither violates deontological prohibitions and obligations nor significantly worsens the distribution of benefits." (Otsuka 1991, 133; emphasis in original)

<sup>42</sup> Otsuka 1991, 135

<sup>43</sup> Ibid., 146 (He refers to Regan 1980)

does not mean that nothing has been added. If another four people were to add another gram each, the scales would flip up to the next 5g point when the fourth one did so, but that would not mean that *nothing* had been added up to that point. It is the same with drops of water. A certain number may have to be added before the person *notices* a change, but that does not mean that nothing has been added (that there has been no benefit) up to that point.”

The problem with this argument (and with Otsuka’s reliance on W) is that it confuses the notion of a benefit with what I will call a benefit *ingredient* – that is, something which contributes towards a benefit.

To clarify, consider occasions on which a goal or benefit is “all-or-nothing”. Suppose, borrowing from Parfit, four men need to stand on a platform in order to operate a winch that will save the lives of a number of trapped miners.<sup>44</sup> There is no benefit unless  $n$  rather than  $n-1$  (four rather than three) contribute. Thus:

$$B_{(n-1)} = B_{(n-2)} = \dots B_1 = 0$$

$$\text{But } B_n = x$$

where  $x$  is the total benefit, i.e. the saving of the lives, and  $n=4$

F is satisfied. But that does not mean that each of the four have not *contributed* to the overall result. One might say that each has contributed one quarter of the *benefit ingredients* (in this case the weight on the platform), but there is no benefit (to the miners) until all have contributed. Thus:

$$BI_1 = \frac{1}{4}(BI_4)$$

$$BI_2 = \frac{1}{2}(BI_4)$$

$$BI_3 = \frac{3}{4}(BI_4)$$

Where  $BI_4$  is necessary and sufficient to bring about  $B_4$

And so:

$$BI_2 - BI_1 = BI_3 - BI_2 = BI_4 - BI_3 = \frac{1}{4}(BI_4)$$

And so:

$$(BI_4 - BI_3) + (BI_3 - BI_2) + (BI_2 - BI_1) = BI_4$$

---

<sup>44</sup> See Parfit 1984, 67-70

Next, consider an intermediate case. Our thousand altruists are building a construction in the garden of a disabled man, to improve his view across the countryside. However, they only have children's Duplo bricks, one each. Each layer of this construction requires 50 bricks, and the man's wheelchair cannot rest on it safely unless a layer is complete. Let us say that the man's view improves by  $x$  amount with the completion of each layer. The following is then the case.

$$B_1 = B_2 = \dots = B_{49} = 0$$

$$B_{50} = B_{51} = \dots = B_{99} = x$$

$$B_{100} = B_{101} = \dots = B_{149} = 2x$$

*And so on until*

$$B_{1000} = 20x$$

But if we consider the benefit ingredients (in this case the bricks).

$$BI_1 = y/50$$

$$BI_2 = y/25$$

*And so on until*

$$BI_{1000} = 20y$$

Where  $y$  amount of benefit ingredient is necessary and sufficient to bring about  $x$  amount of benefit

Finally, return to the original case of the wounded men. It seems that there must be *some* point at which a wounded man feels the benefit of the extra water. (He notices, at the first such point, that he has had *some* water.) The extra contributions up to the point at which he noticed this were additional benefit ingredients (thousandths of a pint), but it wasn't until enough of them had been added that the extra benefit clocked in.

The obvious objection is that it doesn't work like that: there are no markings on the individual "pain measure" as there are on weighing scales. I have sympathy with this view. However, imposing calibrations on a continuum is something that is done all the time. It is something we have to do in order to make much of scientific progress and day-to-day life possible at all. Perhaps the

same is true of this kind of moral philosophy. Moreover, the impossibility of identifying the *exact point* at which something changes is a common one in metaphysics. (When does a human body decay to the point of ceasing to exist? How many planks are replaced in Theseus' ship before it becomes a different ship?<sup>45</sup>) Practicality may require a certain degree of arbitrariness.<sup>46</sup>

A second, related, objection is as follows: suppose we have identified the crucial thousandth of a pint which tips the balance and moves the wounded man up to the next level of pain relief. Then, surely, we could change the example so that that drop could not be identified? If we changed the numbers so that there were two thousand altruists, two thousand wounded men, then one altruist's contribution would only go halfway to filling the gap. If we focus instead on the next drop, then we just have to increase the numbers again, and so on, *ad infinitum*. To this I respond as follows: in theory, yes, but in practice we are limited, ultimately, by facts: for one, by the number of people in the world. (We can only keep on adding altruists up to a certain limit.)

Where, then, does this leave our rational altruist? His extra drop to each individual might make no difference, but it *might* trigger the next stage of thirst reduction (acting in the manner of a 50<sup>th</sup>, 100<sup>th</sup>, or so on, brick in the intermediate case). He knows:

- a) *if* he is the pivotal provider of a drop triggering the next level of benefit, his contributing will have made a difference to each wounded man; and
- b) *if* all thousand rational altruists all contribute, the wounded men will all have their thirst reduced significantly by being able to drink a pint each.

In this case, a truly rational individual should be motivated to contribute. This can be demonstrated by appeal to another argument from Parfit. Ignoring small chances is one of what Parfit dubs mistakes in moral mathematics.<sup>47</sup> As he points out, in calculating the *expected benefit* of an act, the agent has to consider not

---

<sup>45</sup> See Wiggins 2001, 161-71

<sup>46</sup> Consider, for example, the claim that it is mistaken to impose an age limit on driving because one driver of 16 might be better than another of 18: the point is that learning to drive at ten would be unsafe, and at 25 should be safer. Of course there is a degree of arbitrariness in *exactly* where in between the line is drawn, but that does not mean it should not be drawn at all.

<sup>47</sup> Parfit 1984, 73-5

only the chance of its making a difference and the costs of acting, but also the size of the difference.

To use Parfit's example, suppose his vote has a one-in-a-hundred-million chance of affecting the result of an election between Superior and Inferior, the two US presidential candidates. Then, he argues, the expected benefit of voting is:  $(A/B - C)$  where  $A$  is the average net benefit to Americans of Superior's rather than Inferior's election multiplied by the number of Americans,  $B$  is one hundred million and  $C$  is the cost of his voting. 'Given,' says Parfit, 'that there are two hundred million Americans, this sum is likely to be positive.'<sup>48</sup> He adds:

It may be objected that it is *irrational* to consider very tiny chances. When our acts cannot affect more than a few people, this may be so. But this is because the stakes are comparatively low. Consider the risks of causing accidental death. It may be irrational to give any thought to a one-in-million chance of killing one person. But, if I was a nuclear engineer, would I be irrational to give any thought to the same chance of killing a million people? This is not what most of us believe.<sup>49</sup>

The same point can be made here. Suppose that each agent deciding whether to add a brick to the platform for the disabled man does not know how the others will act. He does not know how many of the other 999 altruists will contribute. He knows that out of every 50 "slots" he could fill, one will be pivotal (will push the benefit to the individual up by  $x$ ). He assumes, then, that he has a  $1/50$  chance of increasing the benefit to the individual by  $x$ , which gives an expected benefit of  $x/50$  less the cost to the individual. Given that we have been assuming throughout that the overall project would be considered worth the group's effort, that is, that

$$20x - 1000c > 0$$

(where  $c$  is the average cost to each individual of participating)

then we can assume that, so long as the cost to each individual is roughly the same, that:

---

<sup>48</sup> Ibid., 74

<sup>49</sup> Ibid., 74

$$x/50 - c_i > 0$$

for each individual  $i$ .<sup>50</sup>

An exactly analogous argument can then be made for the altruists and the wounded men, suggesting that rational individuals in the rational altruist case as drawn up by Parfit *should* be motivated to contribute, on consequentialist grounds, even in the case of imperceptible benefits.

However, the discussion does not end there. Firstly, there is a large gap between real human beings and perfectly rational agents, which is likely to be larger when perfect rationality requires making the kind of complex calculations suggested here. How often do we really calculate expected net benefits in making decisions on whether to contribute? Recall Parfit's example. In countries such as the UK where voting is not mandatory, many people simply don't bother to vote, and at least a significant proportion of those decide not to do so on the grounds that it won't make any difference whether or not they do.

It is plausible (and understandable) that each individual would, in the absence of any collectively-led process to reassure her that her action was part of some overall helpful end, decide not to do something which, in itself, appears to make no difference. Thus these scenarios fall under what I call the Rational Altruist Argument A for collective decision making.

Secondly, the argument stumbles over what might be called the overdetermination case. The above calculation of probabilities worked on the basis that, at certain intervals up to the total number of possible contributors, improvements would continue to be made. Suppose, however, that there are 2000 altruists in the case of the platform for the disabled man, but his view will not improve any more after the 20<sup>th</sup> layer has been added. Then, assuming once again that he can only assume that there is an equal chance of each of any number of altruists from 1-2000 contributing, each altruist reasons that his chance of being pivotal is now 20 in 2000, or 1/100. His expected utility falls to  $x/100$  and there is no longer a guarantee that the expected utility of contributing will be positive.

---

<sup>50</sup> The reasoning goes as follows:  $20x - 1000c > 0$  IFF  $20x > 1000c$  IFF  $x/50 > c$  IFF  $x/50 - c > 0$ . But the assumption that the cost to each individual is more or less equal amounts to the assumption that  $c_i$  is approximately equal to  $c$  for all  $i$ .

This leaves us with cases where, it seems, there really is a rational motivation even for an altruist *not* to contribute. Of course, my drawing up of the situation is not necessarily intuitively appealing. Perhaps some readers would be prepared to bite the bullet on the intransitivity implications, and claim, instead, that an imperceptible benefit is no benefit at all. But we have already seen from Otsuka's arguments in the previous section that Parfit's attempt to defend the rationality of contributing would then fail, given overdetermination and the (plausible) assumption that there is some other worthy cause to which each individual could contribute instead.

Can a rational motivation be identified even where an individual knows the chances of making a difference are low, or, given enough other contributions, nil? The next section will engage with an argument by G A Cohen which suggests that individuals should be motivated to contribute even then. However, it will agree only partially, leaving it uncertain whether there will be such motivation, thus grounding Rational Altruist Argument B for collective decision making.

### **VIII. Rational Altruist Argument B: only a chance of benefiting**

Cohen argues that individual members of the proletariat should be motivated to participate in a socialist revolution, *even though* the difference each of them will make is nil. Before considering this argument in detail, it is important to stress its differences from most altruist cases. On the Cohen-Marx account, because workers know that the revolution is inevitable, they *must* know that there will be enough proletariat revolutionaries to make it happen. Thus, says Cohen, an individual worker cannot justify participating *in case* it turns out that he is pivotal. However, as we shall see, Cohen's argument remains useful even in cases without this assumed inevitability.

Cohen argues that each revolutionary participates so that '*he can thereby reduce the burden on other revolutionaries in the task of achieving what they all seek*'.<sup>51</sup> It is important to distinguish this reason from the more general one of reducing the burden on other people in general: this remains specific to the

---

<sup>51</sup> G A Cohen 1988, 63

revolutionary goal. He uses the analogy of a broken-down car with a driver and two passengers. All want the car to be moved. All know this. 'Then,' says Cohen, 'each can consistently believe both that it is inevitable that the car will move and that it is rational for him to join in pushing it: each appropriately motivated person is pivotal here.'<sup>52</sup>

But now suppose there are four people, each resolved that they will push if two others do. What reason does each individual have to push when he knows that there are three others, and only three are needed? Cohen argues:

If an individual pushes, then although, *ex hypothesi*, he does not raise the probability that the car will move, he reduces the burden on each of the other pushers. This suggests a good reason for pushing it, namely, that *each pusher wants to reduce the burden on the others in achieving what they all seek, by assuming some of it himself*.<sup>53</sup>

He draws a distinction, moreover, between the fourth man in this case and the passer-by who stops to help, motivated entirely by an altruistic desire to reduce the burden on the pushers.

Although I am not certain that Cohen's argument convinces in the revolutionary case, he touches on some points important for the rational altruist dilemma.<sup>54</sup> Firstly, *if* one is part of a group and aware of that and aware of shared goals, it is not entirely implausible that one would care enough about the others (or have a sufficient sense of what was right) not to want them to carry too much of the burden. Naturally, this will not transfer to all possible situations, in particular to those where having more people than necessary actively participating in bringing about a goal would *not* reduce the burden on the others already participating, perhaps because what is needed is a particular, irreducible action. Suppose, for example, that instead of pushing the car, the three people had to stand on a platform which acted as a counterweight and lifted it out of a ditch. The fourth person's contribution makes no difference at all.

---

<sup>52</sup> Ibid., 61

<sup>53</sup> Ibid., 62

<sup>54</sup> It is unclear that the second car case *is* an adequate analogy for the revolutionary case. If all four know the situation that they are in, they know that each will be motivated to free ride. Under those circumstances, can Cohen claim that the moving of the car is *inevitable*? I don't think he can. Given that, individuals might be motivated to act for a number of related reasons: to break the deadlock; to increase the chance that the car will move; and to motivate others.



Secondly, Cohen brings out the importance of being a part, oneself, in the achievement of something that matters to one. This comes across to some extent in Parfit's appeal to reluctance to free ride as a potential moral/psychological solution to the problem of collectively self-defeating theories, as well as in his distinction, in discussion of parents' dilemmas, between good being done for one's children and that good being done for them *by oneself*.<sup>55</sup> In the same way, perhaps it matters to an altruist not just that the wounded men are helped, but that he is involved in helping them. This is no guarantee, but it does provide a plausible possible motivation. In many cases, participation might be regarded as an important way of expressing solidarity.

Thirdly, importantly, it may be an oversimplification to treat my participation as adding just one person to the list of participants. In choosing to participate I may motivate others to do so. Again, this will not always be the case, but it may sometimes be.

Thus, in the second component of the Rational Altruist Argument, situations have been identified in which, although an individual *might* have rational motivation to contribute, she would not necessarily do so.

## **IX. Rational Altruist Arguments A&B: changing the unit of agency**

The previous sections left us with grounds for doubting that, so long as individuals are the units of agency, altruist goals requiring the contribution of many will be achieved, even where there is genuine commitment by each individual to the goal. On the one hand, if (as in overdetermination cases) the achievement of the goal has, at least sometimes, to depend on what an individual only *might* be rationally motivated to do, then its achievement becomes doubtful. On the other, even where there is a rational motivation, if the reasoning behind that is complicated enough, and the intuitive impression of the situation is that an individual contribution will "add nothing", individuals cannot be relied on to act entirely rationally. This is especially true if they have no guarantee that others will act, and if their contribution would be wasted if the others didn't.

---

<sup>55</sup> See Parfit 1984: 64-5 and 96-8

These motivation problems could be overcome by a process of co-operation, in which the altruists decided together on the appropriate action and allocated it accordingly. Individuals would know that their individual action was part of a primarily *collective* act. They would no longer have to take chances on being the pivotal person, since the collective enterprise would be organised so as to ensure that the necessary level of contribution was reached. The danger could be removed that, because individuals, *qua* individuals might not be motivated to contribute, the end desired by all would be undermined.

As Susan Hurley argues, if the outcome would be better (the common goal better served) with the collectivity or group as the unit of agency for certain decisions, then it is (collectively) *irrational* not to treat it as such.<sup>56</sup> She claims that arguments such as Parfit's, Otsuka's and Cohen's make a mistake: by taking the unit of agency to be fixed (assuming that the decisions must be made, independently, by individuals acting *qua* individuals), they neglect the option of acting collectively in the strong sense of submitting to some collective decision-making process.

Rather, individuals genuinely committed to the shared goals should be motivated to defer to a collective decision-making process. (In Cohen's car case, this would be a very informal process: the four would discuss how best to utilise the resources at their disposal – say, whether all four should push, making it easier for each pusher, or one of them should walk to the nearby pub while the others are pushing, and buy them all sandwiches – and act accordingly.)

To make the point, consider Cohen's rejection of what looks, at first sight, like a parallel claim – that is, the possibility that 'what I achieve with others can motivate me even where it would make no difference whether I dropped out'.<sup>57</sup> He restrains himself to citing objections briefly in a footnote, but these seem to assume that the unit of agency is fixed at the individual level, which is precisely what Hurley rejects.

Cohen's first three '[c]apsule statements of objections' are:

---

<sup>56</sup> Hurley 1989, 145-8

<sup>57</sup> G A Cohen 1988, 61

- (I) How should I compare, with a view to deciding what to do, the good effects of my redundant collaboration with others with the good effects I can bring about on my own?
- (II) Don't I needlessly sacrifice production of these latter good effects by joining redundantly with others?
- (III) How, without reasoning according to the differences I can make, should I decide which good-effects-producing group to join?<sup>58</sup>

These may hold (as the discussion of Parfit indicated) but only so long as the unit of agency is assumed to be fixed as the individual. His fourth is:

- (IV) Why should redundantly collaborating with other people make sense when no-one would suppose that redundantly collaborating with natural agencies does so?<sup>59</sup>

However, this use of “redundant” in itself assumes an individual unit of agency. Rather, when faced with Cohen's questions, which are pertinent if considered by an individual *as* an individual, the person should bear in mind the wider scenario.

Suppose I am considering petitioning my MP to demand that a local woodland be protected, and I know that if (and only if) at least 500 of her constituents do this, she will act to save it from developers. The case I face is not one in which I know for sure that 500 other wildlife enthusiasts, separate to me, are already committed to achieving our common goal, and so my effort would definitely be redundant. Rather (let us suppose) there are  $m > 500$  individuals all of whom know that only 500 are needed. Thus each knows that *if* 500 or more others write, any further help would be superfluous and energy could be more usefully used elsewhere, but if less than 500 contributed the goal would not be achieved.

These other individuals also all have the same information as me. Given (that we know there is) a big enough  $m$  and a large enough cost of participating, each of us could well decide not to bother. But we all *know* this. Each of us must know, therefore, that the way to achieve the best possible outcome is to evaluate the situation at the systemic level (in Hurley's terms): treat the collectivity (Woodland Supporters) as the unit of agency and be bound by its decisions as to

---

<sup>58</sup> Ibid., 61

<sup>59</sup> Ibid., 61

how to allocate tasks (in this case probably by something as simple as setting up a chat room through which we coordinated our efforts and kept a note of how many letters had been sent).<sup>60</sup> To regard an individual's participation as *redundant* assumes that it must be treated as an individual act rather than as a constitutive part of a collective one.<sup>61</sup>

Hurley, it is worth noting, also contends that it is irrational of the prisoners in the Prisoner's Dilemma, if they are able to act collectively, not to 'act *as a pair*, and thus do better for themselves'.<sup>62</sup> The distinction between individual and collective rationality also applies here, as was brought out in Section IV. However, there is a key distinction in terms of individual motivation. Because, in the Prisoner's Dilemma, there is a clash of short-term *interests*, there will always be a motivation to cheat. Thus the prisoners will only be motivated to act collectively if that collective decision carries with it powers of enforcement.

## X. Clarification and reassurance

This chapter has attempted to demonstrate that it can be better to act *qua* collectivity in pursuit of common or shared goals, purposes or interests, than for individual members to act independently. However, there are some clarificatory points to be made. I have made a wider claim than Hurley. She is concerned *only* with those who volunteer to co-operate.

---

<sup>60</sup> This assumes, of course, that the effort of so doing doesn't exceed the effort of writing a letter on the off chance.

<sup>61</sup> Note, moreover, that there are *efficiency* grounds for acting *qua* collectivity in such cases. Recall Parfit's miners: say there are ten trapped and it requires four altruists to stand on a platform to save them. Suppose there are five altruists, but suppose also there is one miner trapped somewhere else, who requires one man to save him. If Altruist 5 knows that the other four are already on their way to the platform, it is rational for him to save the extra miner. But if all of the altruists learn about the situation at the same time, and are unable to coordinate, each has to decide whether to go to the platform, or go to save the other miner. Because the loss would be so much greater if there were only three or less on the platform, each is likely to go there. However, the end result would be that there were more than needed on the platform, and the additional miner died. If the five could decide, between them, that four (and which four) would go to the platform, then the maximum benefit (save all miners) would be achieved. Leaving it to individuals leaves open the possibility either that there will be more contributions than necessary or (worse) that there will not be enough. Equally, in cases where the burden *can* be reduced by spreading it across members of the collectivity, then the collectivity, *qua* collectivity, could ensure that this was done.

<sup>62</sup> Hurley 1989, 145. Donald Regan makes a case for collective utilitarianism, which he calls "'co-operative utilitarianism'". (Regan 1980, 124)

We should... ask ourselves how we can contribute to the realisation of the best unit possible in the circumstances. This is not the same thing either as individually doing what is best, given what the others do (or are likely to do), *or* as doing what would be best were the others to do it (whether or not they will); in neither of these cases can one's act be regarded as part of a collective act. Collective action does not calculate consequences from a fixed unit of agency; rather, it involves first identifying those willing to act collectively, and then together doing what's best, given what the non-co-operators do (or are likely to do).<sup>63</sup>

I am assuming a slightly different starting point: that of the individuals within a collectivity or group on my non-intentionalist account. These can, recall, be already assumed to share purposes, goals or ATC interests, in the sense of interests/purposes/goals which can only be realised/reached together. The point is that it is *as a collectivity* (not just as those within it who happen to be motivated to make an effort) that decisions need to be made. These collective decisions will then be handed down to individual members. This raises one immediate potential counterexample, and a more general concern. I will address these in turn before leaving the chapter.

The problem case arises if the process of collective action, in itself, undermines the satisfaction that the individual gains from achieving the goal through which she is part of the collectivity. This might come about because an individual has a goal the achievement of which, under the circumstances, would require co-operation with some particular persons, but her feelings towards those persons are such that she would rather not achieve the goal at all, than achieve it by co-operating with them.

Consider Jane, who lives in a small village and commutes to another town. There is no public transport. She wants to cut down her fuel use so it would suit her to car share to get to work. Derek is the only other person in the village who works in the same town, and has the same goal. Unfortunately, Derek and Jane hate each other. Both would rather not satisfy their goal than satisfy it by sharing a car with each other. A wider scale example would be if two nations bordering one another would benefit greatly by trading (each has resources the other lacks) but their citizens are historically ideologically opposed to one other (say, they

---

<sup>63</sup> Hurley 1989, 146.

have clashing religious views) and their citizens would rather not have the benefit of the diverse resources, than engage in trade with a hated neighbour.

Such examples illustrate a special case: one where it is better in terms of the achievement of the goals or purposes through which the set of individuals appear to constitute a collectivity that they co-operate, but not thereby necessarily “better” for each individual member. She might, to use language more familiar from Chapters Two and Three in relation to interests, have a goal which turns out, under the circumstances, not to be an ATC goal.

So long as the mutual dependence is through such non-fundamental goals or purposes, the response to such problem cases is straightforward: co-operation is no longer “better” for the individuals and so there is no reason to co-operate (to take the collectivity as the unit of agency). Thus, obviously, and in contrast to the cases to be discussed in the next chapter, there is no case for coercion.<sup>64</sup> In other words, upon further examination, it seems that such cases do not constitute politically interesting collectivities.

The situation is rather different with regard to interests. Consider the following two cases. Sets of individuals could, like the Rival Gangs in Chapter Two, be mutually dependent through their need to survive even though they hate each other to the extent that individual members might rather sacrifice their chances of survival than co-operate. Equally, members of some collectivities could, although tied to the others for the satisfaction of similarly fundamental interests, put some religious difference above such interests. Consider nations such as Israel and Palestine, all of whose citizens have a fundamental interest in peaceful cohabitation, but at least some of whose members are so ideologically opposed to the idea of cohabiting with each other, that they continue to fight.

However, it was specified in Chapter Two that an individual could only become a member of a collectivity through an unacknowledged interest if it was an ATC interest, and the individual is not automatically taken to be the final arbiter on what is in her ATC interest. Rather, Chapter Three argued that humans have such a fundamental interest in enjoying the central capabilities on something like Nussbaum’s list that it would not be in an individual’s ATC interest to be without a central capability (or expose it to serious risk), even if

---

<sup>64</sup> Chapter Five, Sections VII and VIII

that individual herself believed her interest in enjoying the capability to be outweighed by her personal feelings or ideological goals.<sup>65</sup> Thus, the capabilities approach provides scope for overriding individual views and claiming that it remains “better” (for the individuals) that certain decisions be made collectively.

On these grounds, a strong case can be made for individual duties to co-operate in such cases. This will be explored further in the next chapter. However, (as will be seen) the argument, especially when it comes to the possibility of coercive enforcement of such duties, is rendered more plausible by appeal not just to the fundamental nature of the *individual’s* own interests, but also to the fact that the fundamental interests of others are at stake.<sup>66</sup>

The more general concern with this chapter’s case for collective decision making is that any talk of the “collective” now almost inevitably conjures up a spectre of totalitarianism. I will spell out now why my arguments should not give rise to any such fears.

Firstly, specified in Section I, there is a clear sense in which the interests or goals of a collectivity derive from the goals or interests of individuals, and this is especially so in the case of my more controversial collectivities. (That is, those characterised by unintentionality or involuntariness by some or all members.) In describing an outcome as “better”, I have clarified that I mean “better” in terms of the achievement of goals or satisfaction of interests of the individuals within the group.

Secondly, crucially, I have not drawn any conclusions about *how* collective decisions are to be made. I am assuming (perhaps naïvely) that a fair, reasonable way of making decisions, as a collectivity, can be found. From the discussion of the discursive dilemma, it seems that a conclusion-driven vote will often be sub-optimal, if a vote on independent premises would yield a different outcome, and that sometimes the expert, factual knowledge required means that no vote would be an ideal decision-making process. Considerations of fairness may also be relevant, in particular in selecting from within a set of “better” outcomes

---

<sup>65</sup> Of course, this claim is complicated by the final objection discussed in Section VI, Chapter Four, but I will return to this point in the next chapter.

<sup>66</sup> Chapter Five, II-IV and VIII. The argument is stronger when at least some of those others do *not* regard some other goal (which rules out co-operation) as overriding their fundamental interests. With the Rival Gangs, the case for co-operation is weaker if they all genuinely hate each other so much that they would rather die than co-operate, than if some would rather compromise their ideological beliefs and find a way of getting along, in order to survive.

achieved by collective rather than by aggregated individual actions in partial conflict cases. This, however, opens up a whole new topic.

Thirdly, many collectivities on my model are voluntary. Individuals can move between them, thereby refusing to be committed to goals they reject. (Suppose a sub-group of the environmental activists discussed in Section II decided it was more important to save local trees than any other wildlife. They could always form a rival group to focus on this end.) This brings the cases closer to those considered by Hurley, where it is only those prepared to co-operate, who do so.<sup>67</sup> However, objectors might (rightly) point out that it is more dangerous that individuals have decisions imposed on them when they do not have the “opt out” clause of leaving the group than when they do. This makes the fourth point all the more crucial.

As I will spell out in Chapter Six, it is *only* those decisions and actions directly related to the pursuit of the common goals, purposes or interests which determine the collectivity that I contend should be made by the collectivity as the unit of agency. (Note, moreover, that such “handing down” of decisions is no more than what is already done by (at least allegedly) liberal governments such as that of the US or the UK.) To borrow Dworkin’s example, an orchestra should not dictate to individual players on their sex lives; the UK government should not tell me what colour socks to wear.<sup>68</sup> This is all the more essential when it comes to involuntary collectivities.

As I have already specified, the process of “handing down” decisions may not be a formal one. While at one end of the scale there are formal rules and punishments, or incentives, at the other collective acts are carried out almost automatically. Two campers at the end of a rainy weekend will co-operate to get packed up and out of the rain as soon as possible. If one starts at one end taking out tent pegs, the other starts at the other. Agreement is unspoken: as with Hume’s example of two men rowing a boat, the conventions are taken so completely for granted that they barely register.<sup>69</sup> In between, there are conventions which are closer to rules, because they come with associated

---

<sup>67</sup> Note that, even with voluntary collectivities, there may be short-term situations in which they cannot “opt out” – for example the Monte Rosa climbers half way up a technical bit of the ascent and roped together.

<sup>68</sup> Dworkin 2000, 225-7. Dworkin himself credits the example to Rawls (1971, 520-9)

<sup>69</sup> Hume 1777, 306-7



“social” punishments for infringement (as anyone who has tried to queue jump in at a country post office, or stand on the wrong side of a London escalator will appreciate).

Again, of course, we are left with the apparently paradoxical situation that those situations where coercive enforcement is needed will be precisely those where membership of the collectivity is involuntary or unintentional. That is, in the Game Theory cases or in those collectivities where some members do not acknowledge shared interests. I appeal again to the points above, and to the arguments of Section I. Moreover, where enforcement is *most* crucial, the potential costs of not co-operating are highest – a point which leads us naturally to the next environmental case, and the discussion of the next chapter.

## **Conclusion**

This chapter has identified six plausible scenarios in which it would be better (in terms of achieving the goals, purposes or interests around which the collectivity is constructed) that certain decisions be made by collectivities, *qua* collectivities. The need to take the collectivity as the unit of agency arises because of: inefficiency; ignorance; the conflict between individual and collective rationality in game theory cases; partial conflict; and motivational difficulties even facing rational altruists.

Having thus built on the model established in Chapter Two, I am in a position to take the remaining steps of this project. Firstly, to defend the claim that at least most of humanity constitutes a collectivity on the back of shared interests in environmental preservation. Secondly, to supplement this by identifying another, heavily overlapping set, including, again, most of humanity, which constitutes a morally salient potential collectivity on the back of collective responsibility for environmental harm. Thirdly, to make a partial case for coercive enforcement of collective decisions in each case. And, finally, to bring these strands together and to defend the claim that certain environmental decisions should (prudentially and morally) be made by a global decision-making body, without thereby claiming that there is any need for a so-called global state.

## Chapter Five

### Environmental Harm and Global Collectivities

Chapter Two defended a non-intentionalist model on which a collectivity is constituted by a set of individuals mutually dependent through some common or shared goal, purpose or all-things-considered (ATC) interest. That is, a goal, ATC interest or purpose that it only makes sense to consider as being pursued together, whether because it is genuinely (acknowledgedly) *shared* or because it can only be achieved, for each, by all. The inclusion among such potentially politically relevant collectivities of those whose members are mutually dependent through certain *interests* was supported by Chapter Three's defence of a capabilities model of human flourishing.

This model was expanded to allow for morally salient potential collectivities. Such sets of individuals are acting collectively in the weak sense identified in Chapter Two, Section VI (their actions combine to produce some predictable effect). They would become a collectivity were they to espouse some related goal (preventing, or at least reducing or mitigating the effect). Moreover, they *ought*, morally speaking, to espouse that goal, because they are collectively responsible for harm in the weak but important sense identified in Chapter One.

Chapter Four argued that achievement of the goal, purpose or interest through which a collectivity is constituted will often require *collective* decision making: leaving it to the individual actions of members is often inefficient and can be counterproductive.

The next two chapters will apply these models to the current environmental situation. Chapter Six will lay out some institutional implications, and respond to various objections. This chapter will do two things. Firstly, it will use the theoretical framework to defend collective (and correlative individual) duties to address the problem of environmental destruction. These duties have a twofold grounding. On the one hand, most humans alive today are part of a collectivity because of fundamental interests threatened by damage to the environment. (Section I) This gives us a prudential incentive to establish and abide by such global decision-making institutions as are necessary to tackle the problem of climate change. (Section II).

On the other hand, we (or at least many of us) have a *moral* duty to establish and abide by such institutions. There are two stages to this argument. Firstly, at least arguably, this duty derives from a combination of considerations of fairness, and the positive moral duties, or duties of humanity, that we owe other human beings. (Sections III and VI) Less controversially, we have them on the grounds of collective responsibility for harm – to future generations and to those (generally the global poor) already feeling the effects of climate change. (Sections V and VI)

In the terminology of Chapter Two, an overlapping (though not identical) collectivity and morally salient potential collectivity contain between them all humans. This motivates a collective incentive (or, in the latter case, duty), with correlative individual duties, to set up such global-level decision-making institutions as are necessary to tackle climate change and other pressing environmental problems.

The second step in this chapter is more tentative. Chapter Four not only identified the need for collective pursuit of certain collective goals, but also indicated that powers of enforcement would be necessary in some cases. However, it is one thing to establish a prudential or moral case for individuals to act in certain ways; it is another to justify coercing them into so acting. Section VII will appeal to a bulwark of liberalism, Mill's Harm Principle, as a first step towards defending the coercive imposition of our negatively grounded environmental duties. Section VIII will sketch a line of argument by which coercive enforcement of the duties defended in Sections II, III and IV might be legitimated.

## **I. Self-interest and the environment**

Does the whole of humanity constitute a collectivity through the goal of environmental preservation?<sup>1</sup> A preliminary argument that it does would go as follows. We all have certain fundamental interests (health, affiliation, life, etc.) which are threatened by severe climate change. Preserving the natural world around us is not something any of us can do alone. It can only be achieved, by all

---

<sup>1</sup> Indeed, a case could be made for extending this collectivity to include future generations, although the lopsided nature of the dependence makes this problematic.

of us, for all of us. Thus we are all mutually dependent though our most fundamental interests, and constitute a collectivity.

There are two objections to this line of argument. One can be dismissed (with qualifications). The other is more problematic: it could exclude from the collectivity some of those whose inclusion, from the point of view of any real hope of policy change, is most vital.

The first objection concerns the mutual dependence clause: is the co-operation of *everyone* (all individuals, all tiny nations) really required? Surely if *most* people, across the major nations, implemented emissions cuts, that would be enough to limit environmental damage.<sup>2</sup>

The response to this, very much as with the national case, is that there are two sides to co-operating: acting positively for change, and not doing something dramatically negative to undermine the satisfaction of common goals or interests. Anyone, anywhere, with the ability to develop nuclear weapons, could undermine the attempts at preservation of human, animal and planet life. Anyone who decided to tip sufficient quantities of toxins into the ocean, or to orchestrate a programme of mass deforestation, could do irreparable damage.

We are, to use an overused analogy, like the occupants of a lifeboat. Even if it doesn't take every single individual to row to get the boat to the shore, it does require that no-one leaps up and down and (deliberately or otherwise) overturns the boat, or drills holes in the hull. It follows clearly from the fact that anyone (or at least any, even smallish, group of people) could undermine the project, that its success requires at least some level of co-operation from everyone.<sup>3</sup>

The second objection concerns the "goal, interest or purpose" clause. An opponent would open his case as follows. The whole of humanity cannot constitute a collectivity through a shared goal of, or common interest in, environmental preservation, because not all individuals do share the goal or

---

<sup>2</sup> Assuming that it is not already too late: that we haven't passed some crucial threshold level. (See Gardiner 2004, 562-3)

<sup>3</sup> This prompts the potential objection: since anyone could undermine the projects of almost any other group of people, doesn't that make all of us part of all collectivities? The response is: no, because of the other necessary element of the mutual dependence (sharing in the goal or the interest). A grumpy householder could undermine the shared goals of the collectivity of children playing cricket on the green outside his house by marching out and taking the ball, but that doesn't make him part of the collectivity, because there is no sense in which he shares in their common goal or interest of playing cricket together. However, some such cases (as we will see with George in the next section) do raise the question of *moral* obligations to co-operate.

interest. How could I convince a rich Westerner that he belongs to such a collectivity – and thus has a prudential incentive to promote and abide by collective attempts to tackle environmental damage – if this individual does not care about the state of the planet or the damage to poorer groups or nations through pollution? Suppose, further, that he would derive considerably less enjoyment from life if he didn't continue to drive ten metres down the road to the shops, use CFCs, and so on.

A first response is that such an individual ought (prudentially) to care, because environmental damage is closer to home than he will acknowledge. For at least those of my generation and younger,<sup>4</sup> the near-depletion of energy supplies, and the reduction of usable land through global warming (combined with overpopulation), to name but two, are not problems only “for future generations” or “the poor”. Rather, we can expect them to hit us later in life.<sup>5</sup> They will threaten not only such controversial capabilities on Nussbaum's list as Other Species, but also such undeniably central interests as Life, Health, Practical Reason, Senses, Imagination and Thought, and Affiliation.<sup>6</sup> Doing something to curtail the problem now, before it requires still greater sacrifice, is in everyone's long-term interest.

---

<sup>4</sup> That is, children of the late 1970s and later.

<sup>5</sup> The IPCC predicts a number of serious consequences over the next century for ‘most land areas’. These include: (*very likely*) increased incidence of heat waves, with reduction in water quality and increased water demand, higher heat-related deaths, loss of wildlife and lower quality of life for those without appropriate housing for warm weather, and (*very likely*) increase in precipitation events, damaging crops and causing soil erosion, increased risk of death, injury, respiratory, infectious and skin diseases, disruption of commerce, settlements and transport, loss of property, and pressures on urban and rural infrastructures. (2007, 53) Within Europe, North American and Australia, the report warns of extensive glacial loss, hugely decreased biodiversity, floods and flash fires, crop challenges and water availability problems. (50-2) The US Global Change Research Programme acknowledges: ‘There will be significant climate-related changes that will affect each one of us. We must begin now to consider our responses, as the actions taken today will affect the quality of life *for us and* future generations.’ (USGCRP 2000, 1; emphasis added) It predicts a 3-5°C average increase in temperature in the US over this century – a rise ‘very likely to be associated with more extreme precipitation and faster evaporation of water, leading to greater frequency of both very wet and very dry conditions.’ (7) Moreover, as it points out: ‘The US is linked to other nations in many ways, and both our vulnerabilities and our potential responses will likely depend in part on impacts and responses in other nations. For example, conflicts or mass migrations resulting from resource limits, health, and environmental stresses in more vulnerable nations could possibly pose challenges for global security and US policy. Effects of climate variability and change on US agriculture will depend critically on changes in agricultural productivity elsewhere, which can shift international patterns of food supply and demand.’ (9) In the event of a nonlinear threshold event, those in rich countries would be even more exposed: Gardiner, drawing on Gagosian (2003), notes that ‘it is the rich countries bordering the North Atlantic that are particularly vulnerable to Conveyor shifts’. (2004, 563)

<sup>6</sup> Nussbaum 2000b, 78-80

Thus, most members of current generations, even those not immediately impacted, *do* have an interest in developing an environmentally sustainable way of life. Their general failure to acknowledge this, and to act accordingly, leaves them in a situation analogous to the common land case discussed in Chapter Four.<sup>7</sup>

Suppose, however, the egoistic westerner (call him George) is late middle aged. He won't, he says, be around to be affected by environmental damage, at least to any significant extent, so how can it possibly be in his interest to help to limit it now? There are three possible responses.

The first cites the emotional investment of current generations in at least the next few generations. Climate change and other environmental problems may not affect this particular man, but they will affect his children and grandchildren. Is it not one of his *goals*, even if it is not in his (selfish) interest, to ensure them a high quality of life? For many, it is. But perhaps George has no children. He might (genuinely) claim to have *no* concern for future generations, or what happens to the planet after he is dead. Having as enjoyable a life as possible, himself, is his only goal.

This leaves us with the second and third responses, both of which assert that even George's fundamental interests are put at risk as a result of environmental damage. The second response points out that, even though he will probably be dead before the worst effects of climate change kick in, he is not invulnerable to the natural disasters already increasingly likely as a result of damage to the world's natural stability.<sup>8</sup> We all face growing risk of floods, hurricanes or tornados. Moreover, following Wolff and De-Shalit, awareness of the increased risk (for, say, a resident of Florida) can itself be considered a disadvantage, even to those who never, as it happens, are hit by a hurricane.<sup>9</sup> To this extent, it can be said to be in everyone's interest to reduce the risk of such natural disasters.

---

<sup>7</sup> See Chapter Four, Section IV

<sup>8</sup> For example: 'There is observational evidence of an increase in intense tropical cyclone activity in the North Atlantic since about 1970, and suggestions of increased intense tropical cyclone activity in some other regions where concerns over data quality are greater.' (IPCC 2007, 20) Intense tropical cyclone activity is '*likely*' to increase, bringing with it an increased risk of death, injury and disease, as well as the potential for disruption of water and food supplies, and loss of property. (Ibid., 53; emphasis in original)

<sup>9</sup> Wolff & De-Shalit 2007, 63-73

The third response appeals to the increased risk of *man-made* disaster as a result of climate change. Recall my defence of the characterisation of the nation state as a collectivity.<sup>10</sup> I argued that all individuals can be assumed to have an interest in living in a safe, peaceful society. This draws on the fundamental interests identified by the capabilities approach in Chapter Three. The idea is that everyone is better off under the protection of a system of justice backed up by constraints, than in an insecure state of nature. Using the terminology of the last chapter, this puts the “game” at a wider level: no longer the zero-sum one of how much to tax the rich to give the poor, but a partial conflict one in line with at least some readings of Hobbes, in which certain ways of organising the state and its institutions might be to the advantage of some individuals rather than others, but overall all will be better off being part of the state than not.<sup>11</sup>

A similar claim could be made at the global level. As fuel and usable land become scarcer, without some acknowledged, authorised international body to address the issue, war is increasingly likely. That threatens everyone’s security. The case here is even more extreme than the Hobbesian state of nature: as I have already pointed out, a few nuclear bombs could destroy the whole planet.<sup>12</sup> If the “game” is, as this implies, not zero or negative sum, but one of partial conflict, then surely a collective decision is, again, called for, as ultimately in everyone’s interest?

This is plausible. However, suppose George fully acknowledges the damage being done to the environment and the associated risk to him of natural or man-made disaster. He concedes, therefore, that he has at least *some* interest, other things being equal, in preserving the environment. But other things are *not* equal, and he regards this interest as negligible besides the more important (to him) goal of continuing to enjoy his fuel-guzzling, high-polluting lifestyle. What is in his interest, he points out, is not necessarily in his ATC interest.

This is an important point, and one to which it is difficult to respond. One approach would be to claim that George is simply wrong, as with the terrorist in Chapter Two, who denied that he was rendered part of the nation state by virtue of shared interests. He is mistaken about what is in his own interest. The model

---

<sup>10</sup> Chapter Four, Sections IV and V. See also Cripps 2005, 30-1

<sup>11</sup> Hampton 1986, 132-50. I will return to this in Chapter Six, Section IV

<sup>12</sup> Chapter Four, Sections IV and V

of human flourishing defended in Chapter Three allowed for this possibility. However, this is a rather different case from that of the terrorist, who denies his own ATC interest in health, security or continued life, to the extent of being prepared to accept their almost certain loss rather than abandon his ideological goals. George does not deny the importance of such central capabilities. Nor is he choosing, with near-certainty, to bring about a global war or natural disaster which would seriously threaten them. Rather, he is living a lifestyle which, if he and others continue to live it, puts some of those capabilities to some degree of risk.

It has already been specified that not just any unacknowledged interest can render an individual part of a collectivity: it has to be an ATC interest.<sup>13</sup> Chapter Three argued that everyone has such a fundamental interest in enjoying (continuing to enjoy) the central human capabilities on something like Nussbaum's list that it could not be in their ATC interest permanently to be without a central capability, or to take a serious risk of so being. But it did concede that it could be in someone's ATC interest to expose even one of these central capabilities to some degree of risk, most probably to secure another one.

Some risk is inevitable in a full life.<sup>14</sup> People make choices all the time to take part in activities which put even life or bodily health at risk. At one end of the scale, walking down the road to the shops, driving to Cornwall on holiday or flying to the US for a conference, involves a risk. At the other, people engage in risky sports, which become central to their way of life.<sup>15</sup> (Pursuit of which is included within another central capability category on Nussbaum's list: Senses, Imagination and Thought.<sup>16</sup>) Some may be ignorant of the risks, or ignore them. Others, however, decide, rationally, that they are worth taking.

George runs the risk, in the case of a severe natural disaster, or war, of losing his health, or even his life. However, it is a relatively low risk. George is late middle-aged. His environmentally-expensive tastes are well established and may form a central part of the way of life he values. Given all this, if he, having

---

<sup>13</sup> Chapter Two, Section X

<sup>14</sup> As was noted in Chapter Three, Section IV

<sup>15</sup> As a rock climber, I can make this point with conviction, although such sports do perhaps complicate the matter as the risk can be an intrinsic part of the enjoyment.

<sup>16</sup> Nussbaum 2000b, 78-9



thought it through, believes that it is *not* in his ATC interest that a collective effort be made (including by him) to address climate change, he might be right.

Thus, this section ends only half satisfactorily. Most of us (and the proportion is ever increasing) do constitute a collectivity through a shared fundamental interest in preserving the environment, which would render it in our individual ATC interest to contribute to a global effort to tackle climate change. However, there are some members of current generations whose central capabilities are *not* at sufficient risk for it to be necessarily in their ATC interest to co-operate in addressing the problem of environmental destruction. This is problematic, especially as such individuals tend to be those whose carbon footprint is highest (rich, middle-aged Westerners). Accordingly, Sections IV, V and VI will tackle the question of why, nonetheless, they *should* (morally) co-operate with a collective (global) attempt to tackle climate change. The next two sections will explore the implications of membership for those who are part of the collectivity.

## **II. A prudential “ought”: collective rationality**

Most humans alive today constitute a collectivity with respect to the goal of environmental preservation. The arguments of Chapter Four make clear the need for us to act collectively. Thus, I contend, each of us ought (a prudential “ought”) to promote the establishment of the necessary collective decision-making institutions, then fulfil the responsibilities handed down to them.

This point will be made in two stages. Firstly, I will fill out the prudential case. Secondly, this will be supplemented by appeal to an already influential idea on the moral-political philosophy borderline: Herbert Hart’s Principle of Fairness.

Chapter Four identified various scenarios in which it is better (in terms of the satisfaction of the goals, purposes or ATC interests holding a collectivity together) that the collectivity itself, rather the members *qua* individuals, should be the unit of agency. All these are relevant to the environmental case.

There are obvious efficiency considerations. For example, at the small scale it is more efficient that energy-saving processes such as recycling are coordinated than that each individual tries to recycle alone. Moreover, there will be many who mistakenly deny their membership of the collectivity. This brings the

Ignorance Argument into play. Consider, for example, a fast-driving, high energy-living young man living in Florida. (Call him Young George.) He has not ruled out having children. Moreover his own capabilities are exposed to serious risk in the long term through climate change. Thus, unlike George, he is a member of the collectivity. Nonetheless, he acts like the Head-in-the-Sand Farmers and ignores or denies such considerations.<sup>17</sup> A more extreme case is the fundamentalist terrorist, who would rather sacrifice his own life and those of the countless others than accept a world characterised by what he sees as “evil” ways of life.

Nor does acknowledging our shared interest, in this sense, give each of us an incentive, *qua* individual, to cut carbon emissions (etc), unless we can rely on others doing so, which we cannot unless this is enforced because even having made an agreement, each of us would be better off cheating. (Game Theory Argument.) Even if we did all adopt the *goal* of tackling climate change, there are various solutions that some individuals would prefer (which benefited them more, or which better suited their ideological views) and various (other) solutions which would be better for other individuals. This gives rise to Partial Conflict problems. Moreover, even without Partial Conflict, each of us would be faced with the motivation problem that our own reduction in carbon footprint would make an imperceptible difference. (Rational Altruist Dilemma.)

The Game Theoretic and Rational Altruist Arguments bring out the distinction, touched on above, between two ways in which an individual could contribute to a collective effort. Individual co-operation could take the passive form of not setting off a nuclear bomb, setting fire to the rain forests, or similar, or the more active form of cutting our own everyday carbon emissions. While all of us (bar the terrorists) would acknowledge a prudentially grounded responsibility to co-operate in the first sense, we do not *all* need to co-operate in the second sense (cut down our carbon emissions) for the overall goal to be achieved. Moreover, this kind of action by each of us individually could be a pointless sacrifice if not enough others act similarly.

This is precisely why our prudentially grounded incentive is to promote a *collective* solution. (This might, albeit awkwardly, be called a “duty to oneself”.)

---

<sup>17</sup> Chapter Four, Section III

Reasoning parallel to Hobbes, each of us has a rational incentive (it is a “law of reason”) to reduce our carbon emissions *so long as we can be confident that enough others will do so too*.<sup>18</sup> Because (a) we are not confident enough about our fellow humans to believe that they will do so, and (b) we know that it would be more efficient to organise the reduction across the collectivity than risk all over sacrificing unnecessarily, the prudential course of action for each of us is to promote the necessary collective decision-making apparatus.<sup>19</sup> This is collectively rational, and satisfies the common good as defined by Sen and Runciman.<sup>20</sup>

### III. A moral ought: the Principle of Fairness

This argument can be bolstered by normative considerations. The question of why we should contribute to collective schemes from which we benefit is a familiar one in political philosophy. There is already a tool available with which to defend the duty of each and every member of the collectivity to do their part in a collective effort to tackle climate change: the Principle of Fairness.<sup>21</sup> That is, Hart’s claim that:

when a number of persons conduct any joint enterprise according to rules and thus restrict their liberty, those who have submitted to these restrictions have a right to a similar submission from those who have benefited by their submission.<sup>22</sup>

---

<sup>18</sup> Hobbes defends his first Fundamental Law of Nature as ‘a precept, or general rule of Reason. *That every man, ought to endeavour Peace, as farre as he has hope of obtaining it; and when he cannot obtain it, that he may seek, and use, all helps, and advantages of Warre.*’ (1651. 91-2)

<sup>19</sup> Note, however, that it also follows from the arguments of Chapter Four, Section IV, that for many of those involved this is not a Rational Altruist dilemma, but a Game Theoretic one, analogous to the Tragedy of the Commons, in which case the “necessary” decision-making apparatus would be one with powers of coercion. I will return to this in Section VIII.

<sup>20</sup> Runciman & Sen 1965, 555-6. See also Chapter Four, Section I

<sup>21</sup> This move is not a new one. As Casal notes: ‘Rawls’ widely discussed *principle of fairness* is often thought to support the provision of environmental public goods.’ (1999. 363) However, I will offer what I hope is a new slant on the discussion, by appealing directly to the idea of our positive moral duties to preserve the fundamental interests of other human beings. (Sections IV and VIII)

<sup>22</sup> Hart 1955. 185. A version of the principle is also famously espoused by Rawls: ‘[A] person is required to do his part as defined by the rules of an institution when two conditions are met: first, the institution is just (or fair), that is, it satisfies the two principles of justice; and second, one has voluntarily accepted the benefits of the arrangement or taken advantage of the opportunities it offers to further one’s interests.’ (1971. 96)

As it stands, this principle is liable to attack. According to Robert Nozick, it is 'objectionable'. He appeals to the examples of a neighbourhood radio station ('Though you benefit from the arrangement, you may know all along that 364 days of entertainment supplied by others will not be worth your giving up *one* day.') and a local rota to sweep streets. ('If each day a different person on your street sweeps the entire street, must you do so when your time comes? Even if you don't care that much about a clean street? Must you imagine dirt as you traverse the street, so as not to benefit as a free rider?') At the very least, he argues, the principle would have to incorporate the 'very strong condition' that the benefits to each individual of co-operating outweigh the burdens, but it would remain objectionable even then, as individuals might prefer some other scheme altogether, or benefit much less from this one than others do.<sup>23</sup>

In response, George Klosko makes a move similar to that made earlier in this thesis, when it was specified that the only *unacknowledged* interests through which individuals could count as members of collectivities were ATC interests.<sup>24</sup> He amends the Principle of Fairness to apply only subject to certain further conditions: 'the goods in question must be (i) worth the recipients' effort in provide them and (ii) "presumptively beneficial"' where "presumptively beneficial" goods are 'something similar to Rawls' primary goods... such goods must be necessary for an acceptable life for all members of the community',<sup>25</sup> and 'the costs and benefits of the [co-operative] schemes in question are fairly distributed'.<sup>26</sup> Call this the Modified Principle of Fairness.

Such presumptively beneficial goods need not, to make Klosko's point or that of this chapter, include anything other than central capabilities. Nor need they even extend to the more controversial capabilities on Nussbaum's list (although an interest in Other Species would certainly add force to the arguments offered here). As Klosko puts it: 'such things as physical security, protection from a hostile environment, and the satisfaction of basic bodily needs appear obviously to fit the bill'.<sup>27</sup>

---

<sup>23</sup> Nozick 1974, 93-5

<sup>24</sup> Chapter Two, Section X

<sup>25</sup> Klosko 1987, 246. Casal draws a similar distinction, between *conditional* public goods, the value of which 'derives entirely from them satisfying individuals' preferences' (1999, 369) and *unconditional* goods, which are supposed to have value over and above those preferences.

<sup>26</sup> Klosko 1987, 253. One of his examples is a society faced with pollution (1987, 250)

<sup>27</sup> Ibid., 247

The Modified Principle of Fairness avoids requiring compliance in Nozick's examples, but adds force to the case made by the previous section. There, it was argued that members of the collectivity (including Young George) had a prudential incentive to commit to a collective scheme to tackle the problems of climate change. Presumably, for that incentive to remain, the individual would have to be confident that the collective decision-making process would divide duties in such a way that the effort required of one, as an individual, would not outweigh the costs – that is, that everyone else contributed their share. Given that each individual would reason thus, the possibility of a workable collective solution would presumably hinge on its ensuring general compliance.

The Principle of Fairness reinforces this with a moral consideration: the *unfairness* of a situation in which all benefit, all things considered, but only some actively contribute. (This might become relevant, in practice, if in the process of establishing the decision-making institution, those in a superior bargaining position, perhaps because less *immediately* at risk from climate change, played a tactical game and refused to co-operate unless they bore a smaller share of the burdens.)

Thus far, then, it has been argued that most of those currently alive, including Young George, constitute a collectivity with respect to the goal of environmental preservation, and that this gives members a prudential reason to promote and comply with an appropriate collective effort to tackle the problem. We have supplemented this with an appeal to the *fairness* of contributing to some collective effort which protects a fundamental interest, where the cost of contributing is less than the benefit.<sup>28</sup>

George, however, represents a more problematic case. If it is not in his ATC interest that a collective effort be made (including effort by him) to tackle

---

<sup>28</sup> A further objection might be raised here: how can the costs not outweigh the benefits for some members if the cost to some people is their life? Klosko acknowledges that it might be necessary in a military campaign to sacrifice some soldiers. 'for the good of the larger unit' but claims that those soldiers would nonetheless be obliged to fulfil that duty, so long as 'the immediate benefit of this unfortunate group's mission is sufficiently large, and the members of the group are chosen fairly'. (1987, 251) How could this be defended? Klosko appeals to the good for the group *as a whole*. However, it might be possible to render even the military case compatible with an individual balance of benefits and burdens, by applying condition (i) *ex ante*. If each individual would be prepared to put her life at a certain degree of risk to secure a certain benefit (not being invaded, say), then it could be argued that a "fair" system (say lottery) to pick who goes on a particular mission would be enough to satisfy the condition.

climate change, why should he be required to contribute to such an effort? The next section will consider this question.

#### IV. Motivating George: duties of humanity

Why should George promote and contribute to collective efforts to tackle climate change? Given that he benefits from environmental preservation (the world is safer for him, as well as for everyone else) one option would be to appeal again to the Principle of Fairness.

However, this prompts Nozickian worries: George would rather that there were no scheme (and no benefit) than that there were one and he had to contribute. Why, then, should he do so? Section III showed how Nozick's counterexamples could be avoided by restricting the principle to cases where (i) the benefits of co-operation outweighed the burdens and (ii) the good was "presumptively beneficial". Such a restriction, however, seems to exclude George, at least so long as we read (i) (as I was implicitly doing in Section III) as requiring that the benefits outweigh the burdens *for each participant*.

George benefits from the scheme, in that a risk to his central capabilities is reduced. In that sense, he shares in a presumptively beneficial good. However, it has already been acknowledged that the removal of *all risk* to such capabilities is not necessarily in his ATC interest. The reduction of an already small risk, at the cost of curtailing his ability to pursue his own plan of life, will, for him, not be worth it.

Klosko avoids this objection by specifying a weak reading of his condition (i): '[T]he benefits and burdens under consideration are those *of the relevant community as a whole*, rather than of each particular member.'<sup>29</sup> (Call this the Weak Modified Principle of Fairness.) This is a significant shift. It also requires further defence, both for the cases with which Klosko is concerned, and for the global environmental case.

On the non-intentionalist model, there is no understanding of the community or collective interest which does not derive directly from the interests of the individual members. In a full blown collectivity, each individual has a

---

<sup>29</sup> Ibid., 250 [emphasis added]

fundamental interest (or acknowledged goal) and is mutually dependent with the other members for its satisfaction. Here, however, the situation is asymmetric. Not everyone has an ATC interest in collective efforts to achieve environmental change although they do benefit to some degree. Thus, the supposed community “good” (or balance of benefit over burden) would have to be understood as that which is in the ATC interest of *most* but not all, and which requires the co-operation of all for its satisfaction. Given that, Nozick’s objection retains its intuitive pull: if participating in some overall scheme will cost *me* more than it gives me, why should *I* comply with it?

Klosko does not provide an entirely satisfactory response to this. He glosses over the distinction between the two cases (overall and individual balance of benefit and burden) by referring to the typical individual: ‘[F]or condition (i) to be satisfied, the benefits must obviously outweigh the burdens (for the community as a whole, and so for the typical individual).’<sup>30</sup> The problem, however, is justifying duties to co-operate for an *actual* individual who does not benefit.

Hart, defending his Principle of Fairness, appeals to ‘the [natural] equal right of all men to be free’.<sup>31</sup> This, he claims, underlies all moral rights, including special rights. Thus, when someone makes a promise, or confers some right on someone else, ‘this claim to interfere with another’s freedom is justified because he has, in exercise of his equal right to be free, freely chosen to create this claim’.<sup>32</sup> However, his attempt to derive the kind of “mutual restriction” rights covered by the Principle of Fairness from our equal right to be free is less straightforward.

[I]n the case of mutual restrictions we are in fact saying that this claim to interfere with another’s freedom is justified because it is fair; and it is fair because only so will there be an equal distribution of restrictions and so of freedom among this group of men.<sup>33</sup>

This involves a step from the “natural” right to freedom, which we all hold equally, to the idea that, when a system is set up which restricts freedom, those

---

<sup>30</sup> Klosko 1987, 252

<sup>31</sup> Hart 1955, 175

<sup>32</sup> Ibid., 190

<sup>33</sup> Ibid., 190-1

restrictions should be evenly spread: that is, from an equal *right* (to be free) to a right to equal *freedoms*. Accordingly, as Klosko notes, modern scholarship has acknowledged the reliance of the Principle of Fairness on another general moral principle: fair divisions of benefits and burdens.<sup>34</sup> Plausible as this when everyone *does* benefit, here it only brings us back to Nozick's point. If, for me, the benefits of taking part in a system of co-operation are *less* than the costs to me of a "fair" distribution of burdens between me and others, within such a system, then surely it is *unfair* to assign me an obligation to participate, at all? This is precisely George's complaint.

Further argument is needed to explain why George should contribute. This, I suggest, can be done by taking a different slant on Klosko's appeal to the good as presumptively beneficial to the community as a whole, or by elaborating on the following point from Casal.

[E]nvironmentalists are primarily concerned with what one may call *unconditional* rather than *conditional* public goods. They think that the value of the unconditional goods they try to secure extends beyond the satisfaction of their own personal preferences. It may derive from the intrinsic value of ecosystems or from the value of individual animal lives. *Or it may depend upon the critical interests of human beings other than those who secure the environmental good at stake...* [O]ne does not need to love the forests in order to have a reason not to burn them.<sup>35</sup>

Crucially, the central capabilities of the *other individuals* involved are at stake. If, without the co-operative system, their most fundamental interests would be exposed to serious risk, we can appeal to George's positive moral duty (to prevent serious harm to others) to defend his duty to co-operate. Note that exactly the same argument can be used to supplement the claim that Young George, or the terrorist, *ought* to contribute even if they deny the shared interests. This is important for Section VIII.

A smaller-scale analogy brings out the intuitively important, and morally relevant distinctions in play here. Compare the following situations (which are not dissimilar to the Monte Rosa examples in Chapter Two).

---

<sup>34</sup> Klosko 1987, 242, citing David Lyons: *Forms and Limits of Utilitarianism* (Oxford University Press, 1965)

<sup>35</sup> Casal 1999, 369 [emphasis added]



- (a) Anne, Beth and Costas have a shared goal. They want to climb a mountain by a particular route. David does not share their goal. They can pursue it perfectly well without him.
- (b) Anne, Beth and Costas share the goal of doing the climb. David does not, but David is the only one who is a technically good enough climber to lead the expedition: they can't do it without him.

Consider also the further distinction between the following:

- (bi) The situation is as in (b) but Anne, Beth, Costas and David are all in a café in Edinburgh. Anne, Beth and Costas might be annoyed at not being able to do the climb, but it will not do them any serious harm.
- (bii) The situation is such as in (b) but the four of them are stranded somewhere remote in the Highlands, where a night in the open would be extremely dangerous, and only David has sufficient experience and gear to be likely to survive it. The climb is the only way out.

Intuitively, it seems that David ought to participate in (bii), but the others can make no such claim on him in (a) or (bi). This can be explained by the combination of two factors: the goal or interests of the others cannot be satisfied without him (as they can in (a)); and (unlike in (bi)) the stakes, for those others, are very high. If David (at relatively low cost to himself) can prevent their deaths, he ought to do it. He has, it might be said, a positive duty, or duty of beneficence, to help them, and on these grounds ought to adopt the relevant goal and become part of the collectivity.

Put more generally, there is an important difference between saying individuals ought to act as if they were part of a collectivity because of mutual dependence for the satisfaction of interests which they do not acknowledge, and which may not be in their ATC interest: (i) when it is possible for the collectivity to exist and to pursue those goals/interests successfully without them, and when it is not, and (ii) when the failure to achieve the goals constitutes a threat to a fundamental interest (or central capability) of the individual members, and when it does not.

With this in mind, reconsider the case of the whole of humanity in relation to environmental preservation. Given that the interests of the relevant others (including younger generations in the west, as well as those in poorer areas already negatively impacted by climate change) are so fundamental, and given

that the co-operation of the Georges of the world is necessary to achieve the goal, the case seems comparable to (biv). Refusal to co-operate would thus constitute a moral failure: a failure to protect the fundamental interests of others.

However, the analogy is not so straightforward. It is true that everyone's co-operation is required in the sense outlined in Section I: that they refrain from setting off nuclear bombs, emptying tankers of toxins into the sea, and so forth. (Moreover, their responsibility to co-operate at least thus far can be grounded, less controversially, in the *negative* moral duty not to harm others.) However, when it comes to the more focused attempts to cut carbon emissions, George can point out that his continuing to drive everywhere, overheat his house, and so on, would not prevent the achievement of the overall goal, so long as everyone else co-operated. His *individual* refusal would not undermine the whole project, so how can we appeal to his positive moral duties (to prevent harm to others) to require that he contribute?

We can do so, I suggest, by appeal to the concept of collective responsibility. While his individual co-operation (in this more positive sense) may not be necessary, the co-operation of *people like him* ("the Georges of the world", as I said above) *is* necessary. If all middle-aged westerners reasoned as he does, the efforts of the rest of the world would be inadequate. Thus, while George himself is not directly analogous to David from the climbing example, he and those like him are *collectively* so.

Chapter One made the case for taking collective moral responsibility seriously. Now, however, it is not even necessary to rest on the extension of May's theory defended there, as May himself provides a model of collective responsibility for collective *inaction*. As discussed in Chapter One, Section II, this refers to 'the failure to act of a collection of people that did not choose *as a group* to remain inactive but that could have acted as a group'.<sup>36</sup> Thus, May argues, the holidaymakers on a beach who could have organised in time to save a drowning child, but have not done so, are collectively responsible for failing to prevent the death of the child.

Similarly, if George and those like him do not, between them, do something to reduce *their* carbon emissions significantly, they will be failing to prevent serious

---

<sup>36</sup> May 1992, 107

harm. As with David, it seems that they, collectively, have a positive responsibility to co-operate, and George, in parallel with the arguments of Section V, Chapter One, has an individual duty to promote the fulfilment of this collective duty.

This can be tied back into the discussion of Murphy in Chapter One, Section V, the point being that it would be morally suboptimal if each of “the Georges of the world” did not contribute, as the sacrifice required of those who did would be correspondingly unfairly greater. Indeed, according to Murphy’s Compliance Condition, each could only be morally *required* to do what would be his share under full compliance.<sup>37</sup>

This section has attempted, by supplementing the prudential with a moral argument, to show that even George ought to espouse the collective goal of tackling climate change. However, any argument resting on *positive* duties, or duties to aid others, is inevitably open to dispute. Such duties are not only less clear cut than their negative counterparts, but are also often taken to be weaker and so liable to being “outweighed” either by individual considerations or other, negative claims.

Philosophically, it is unclear how much weight should be given to the negative-positive distinction.<sup>38</sup> (For example, if I shut the door of my isolated home on a stranger on a snowy night in the Canadian winter, am I failing to help or harming him?) Nonetheless, it is deeply embedded in the common sense morality guiding our everyday lives.<sup>39</sup> Moreover, there is greater controversy regarding the legitimacy of enforcing positive duties than negative ones. Comparing the discussion of the Harm Principle (Section VII) with the return to the Principle of Fairness in Section VIII will bring this out.

It would, therefore, be an advantage if the defence of environmental duties did not rest on positive moral responsibility alone. Happily, thanks to the theoretical apparatus already built up in this thesis, it need not. The morally salient potential collectivity constituted by most of us alive today (including George) incurs a *negatively-grounded* collective duty, which comes with correlative individual

---

<sup>37</sup> Murphy 1993, 278

<sup>38</sup> See, for example, LeRoy Smith (1978) and Foot (1978a)

<sup>39</sup> Scheffler makes this point (2001, 36-8)

duties, to address the problem of environmental damage. The next section will defend this claim.

## V. Environmental harm and moral excuses

Chapter One defended a notion of weak collective responsibility for the harm which results from the predictable aggregation of individual acts.

One such case is damage to the environment, which occurs through the aggregation of the acts of individuals (as well as other agents, such as corporates and governments). The resulting harm impacts, uncontroversially, on two sets of individuals, to the extent of undermining their central capabilities: those (generally the poor) whose chances of a flourishing life are already undermined by factors such as pollution and global warming;<sup>40</sup> and future generations (including the current generation of children).

Accordingly, “we” (the “global elite” constituted by westerners and the rich minority in poor countries<sup>41</sup>) are collectively, morally, responsible for harm. This renders us a morally salient potential collectivity: one which ought to count itself *as* a collectivity, with respect to the goal of curtailing environmental destruction. We have a collective duty to end (or at least reduce or mitigate) the harm.<sup>42</sup> Individually, we have correlative duties to promote a collective solution (in this case, in the first instance, a collective decision-making body) and possibly supplementary duties as well.

However, before drawing this conclusion, along with all its implications in terms of collective and individual duties, it is necessary to consider a potential line of objection. Certain excuses are standardly used to argue that an individual who is responsible for some harm, has nonetheless acted *permissibly*. The harm is not morally *unacceptable*. Is it possible to appeal to collective equivalents of

---

<sup>40</sup> Shrader-Frechette 2002, 6-13

<sup>41</sup> A grouping identified by Pogge in the global justice debate. The question, he says, ‘is not what we are doing to the developing countries? The crucial question is what are we and the rules and elites of the developing countries *together* doing to their impoverished populations?’ (2002, 23)

<sup>42</sup> Lomborg has argued that it would be more cost effective to adopt a policy of adaptation to climate change rather than to attempt to prevent it. However, as Gardiner has pointed out, this rests on a high discount rate, so that the impact to future generations does not count as a significant cost. Even if Lomborg were correct, however, we would still have strong environmental duties: duties to enable adaptation to climate change. (Lomborg 2001, 317; Gardiner 2004, 569-73)

these excuses to deny (or override) any moral duty, collectively, to address the problem of environmental harm?

Consider the following four scenarios, at the individual level, in which the individual's causing the harm might be morally "excused":<sup>43</sup>

- 1) *Unacceptable cost to self*: the cost to the individual concerned, of not bringing about the harm, is too high.
- 2) *Worse alternatives*: in not causing this harm, the individual would cause (or fail to prevent) some other, far greater harm.<sup>44</sup>
- 3) *Consent*: the person (or persons) being harmed has (have) given her (their) free informed consent to the harm.
- 4) *Acting within one's rights*: although an individual is causing harm to some other person or persons, she is acting within her rights in so doing, and so they have no claim against her.

The first three can be "collectivised", and considered as potential excuses to which "we", the global elite, might appeal. The fourth is different. It has most force considered as an excuse, once collective decision-making institutions are in place, for an individual not to comply with the corresponding restrictions on individual behaviour. It will thus be saved for the next section.

There is also a fifth possible excuse.

- 5) *No evidence*: there is insufficient or inconclusive evidence that action by the individual is bringing about the harm.

---

<sup>43</sup> I am aware that there is a distinction, in legal philosophy, between "excuse" and "justification". 'Justified behaviour,' says Robinson, 'is correct behaviour and therefore is not only tolerated but encouraged. In determining whether given conduct is justified, the focus is upon the *act*, not the *actor*... Excuse, on the other hand, focuses on the actor, rather than the act. A successful defence of excuse represents a legal conclusion that although the act was wrong, liability is inappropriate because some characteristic of the actor vitiates society's desire to punish him.' (1975, 289. The article is reproduced from *UCLA Law Review* 266) See also Austin 1956 and Husak 2005. However, with the *moral* distinction itself a matter for debate, and nothing hinging on it as far as this work is concerned, I will understand moral "excuses" as also including justifications.

<sup>44</sup> It might be argued that (1) is a subsidiary of (2) - harm to *oneself* being a variant of harm to others. However, I see them as importantly different. What you might be expected to bear yourself, rather than pass it on to others by harming them, might very well be more than what you can legitimately do to one person, in order not to harm another.

This bears mentioning, but the collective equivalent in the environmental case (though frequently invoked) can safely be dismissed. The scientific evidence is overwhelming that human activity is causing, or at least dramatically worsening global warming. For example, a summary of the scientific literature by Naomi Oreskes clarifies that, contrary to the claims of various corporations and interest groups, ‘there is scientific consensus on the reality of anthropogenic climate change’.<sup>45</sup>

Consider, then, (1) to (3) in turn. Firstly, unacceptable cost to self. Generally, the force of this excuse depends not on the absolute cost to oneself of not harming, but on the cost *compared with* the cost of the harm to the person harmed. For example, jumping out of the way of a speeding car to save one’s own life, and in the process (knowingly) knocking over a pedestrian, would harm the pedestrian to the extent of a few bruises. However, it would not standardly, or intuitively, be considered that one would therefore have a duty not to do it. Where the consequences to oneself of not harming are less bad, harming is less acceptable. Thus, it would seem that there is a moral duty *not* to knock over the pedestrian by running along a narrow pavement to avoid being late for work. Equally, the worse the harm done to the other person or persons, the less the chance that even the avoidance of a significant cost to oneself renders it acceptable. (Pushing another person in front of the car in order to save one’s own life, for example.)

Applying this collectively, recall the example of teenagers jumping independently into the lake, and between them bringing about the death of a child.<sup>46</sup> The teenagers, it was argued, have a collective duty to adapt their behaviour so as not to cause the drowning of the child. It is hard to think of any scenario in which the cost to them could eradicate this duty. However, changing the circumstances, if the harm to the child would be mild discomfort rather than

---

<sup>45</sup> Oreskes 2004, 1686. A more plausible variant on (5) might be offered, however: that there is insufficient evidence as to what is to be done to remedy the harm, and even as to whether it is possible. Neither renders blameless our continuing to harm. In the first case, extensive research into possible approaches would be a first step towards fulfilling the collective duty. In the second, lack of certainty that disastrous consequences can be averted is no excuse not to try – and certainly no excuse to carry on harming – given that the stakes are so high. (Although, were there incontrovertible evidence that nothing can be done, then a different set of responsibilities would arise, relating to bearing the consequences with as little undermining of flourishing as possible.)

<sup>46</sup> Chapter One, Section III

drowning, and the teenagers were freezing to death outside a (heated) pool, this consideration would presumably override the duty.

In the environmental case, any such claim seems wildly implausible, at least as regards the global elite.<sup>47</sup> Much of our greenhouse gas output serves luxury purposes.<sup>48</sup> There is one scenario in which we might claim to need to harm other generations in order to preserve ourselves from at least as bad harm. That is, in which the only way to protect ourselves from potentially huge suffering as a result of climate change would be by storing up still greater damage for future generations. However, there are two points to be made about this.

Firstly, at the individual level, it is one thing to say that I would be morally permitted not to save your life if the only way I could do so would be by losing mine. It is another to claim that, if it were a choice of my life or yours, I would be morally permitted to kill you to save myself. It is still another to claim that, in a situation which is at least in part of my creation, and not at all of yours, where I could save myself by sacrificing you, that I would be morally permitted to do so. The parallel is, I hope, obvious.

Of course, the situation is complicated by the fact that current generations are not the only ones doing the damage: much was done before we were born, or at least before we were adults.<sup>49</sup> Were some future generation (which was not itself over-emitting) to face the stark choice between saving itself or saving some distant future generation, there could be a genuine moral dilemma. However (the second point), that is not the situation now. We are not innocent victims faced, though no fault of our own, with the choice of harming ourselves or others severely. The situation is at least in part of our making, and we are not yet (at least in the west) facing real suffering as a result of it.

Consider, then, the second get-out clause: *worse alternatives*. It is difficult to see what worse harm to others we could be supposed to be preventing by

---

<sup>47</sup> As Shue points out, in the slightly different context of defending redistribution to poorer countries to help accommodate the costs of *coping* with climate change, the interests at stake for those in poorer countries are 'vital interests—survival interests. This means, I think, that it is unfair to demand that they be sacrificed in order to avoid our sacrificing interests that are not only not vital but trivial.' (1992, 394)

<sup>48</sup> Shue emphasises this distinction (1993, 54-8). The terminology is from Agarwal and Narain (1991, 5-7). For example, they cite the International Project for Sustainable Energy Paths to argue that a 50 per cent per capita reduction in beef consumption in industrialised countries, which would still allow a healthy diet, would reduce animal methane production by 40 per cent.

<sup>49</sup> I will come back to this point in Chapter Six, Section VI

continuing, collectively, to do environmental harm. One suggestion would be that we should focus on today's global poor, rather than future generations. There are various responses to this, none of which deny the urgent need to tackle poverty, and the claims (undoubtedly of humanity, arguably of justice) that the global poor have on the wealthy elite.

The global poor (or at least a significant subset of them) are already harmed by our emissions levels. While it is sometimes plausible that harming someone can be in their ATC interest (that is, if the harm is a side-effect of some activity which gives them a greater benefit), a strong empirical argument would be needed to show that the benefits even to this generation of the global poor are higher than the costs. (Rather, it seems, *we* reap most of the benefits and *they* bear the bulk of the costs.) Even if there were such an argument, refusing to tackle climate change would simply shore up greater problems for the poor in future generations. So long as climate change gets worse, poverty can be expected to do the same.

Moreover, we are not failing to tackle climate change *in order* to tackle global poverty. We fail to cut emissions levels because we choose to continue to drive cars to the shops, fly abroad for holidays, leave TVs on standby, turn the heating up rather than wear a sweater, and so on.

The third excuse, *Consent*, has to be evaluated with regard to both harmed groups: future generations and those in current generations already suffering the effects of environmental damage. To begin with the former, it seems immediately obvious that they not only could not, in practice, but would not, in principle, consent to being left with a world damaged, to the extent of severely limiting their health and their ability to flourish.

The point can be made with an analogy. Suppose my neighbour says to me: "I want to have a huge party, with a giant bonfire, as a result of which your garden will be damaged by smoke, and rubbish will be piled up in it. This puts yours and your family's health at risk. You are not invited to the party, so you will get no enjoyment from it. Do you consent?" Unless I am altruistic to the



point of masochism, I will not. There is no reason to suppose that future generations are this altruistic.<sup>50</sup>

This also undermines the suggestion that future generations can be considered to have consented because members of current generations, acting as their representatives, have done so. Kristin Shrader-Frechette points out:

Such second-party consent is at least *prima facie* plausible because, in a democracy, we recognise that all citizens must make some concessions to one another, to majority rule, to operate a constitutional government. At least one... could be that our representatives can make decisions for us and therefore that the representatives of future generations could make decisions for them.<sup>51</sup>

However, in a democracy, those on whose behalf decisions are made play a role in appointing the representatives. This is not the case with current governments deciding for future generations. Moreover, it is not part of the idea(l) of democracy that those making the decisions make those best for *them*, or for a subset of the overall set represented, at the expense of other members. Rather, the decision is supposed to represent the interests of the whole set – to be the decision that they, with the proper information, would have made. This is clearly not the case here.

This leaves us with the second element of the consent excuse: (alleged) consent by current generations. This excuse is familiar from cases where there is harm to a specific individual or group of individuals: the African worker, it is suggested, *consents* to take a job with enormous health risks; the Native American or poor African American community *chooses* to allow waste to be dumped on its land, in return for payment; poor countries *consent* to take waste from the West. Such cases have been extensively discussed by Shrader-Frechette.<sup>52</sup>

There are three responses to this. The first is to note that such cases generally refer to specific acts of environmental damage, usually where there is at least

---

<sup>50</sup> I refer the reader unconvinced by my brief treatment here to Shrader-Frechette's detailed discussion of the topic in relation to harm to future generations through (allegedly) permanent geological disposal of nuclear waste. (2002, 105-13) She also dismisses a consent excuse in the case of a waste incinerator imposed on an African American community. (77-84)

<sup>51</sup> Shrader-Frechette 2002, 106

<sup>52</sup> *Ibid.*, 71-93 and 124-9

some notional payment to those (or alleged representatives of those) accepting the harm. It is (even) less plausible to claim that those harmed have consented (or would consent) when it comes to the harm done to those (generally the global poor) feeling the impact of pollution and climate change on the planet as a whole.

Returning, however, to the narrower cases, the second response is that the claim that there has been free informed consent often falls apart upon investigation. The third is that cases where individuals *do* consent to being exploited only serve to show that consent, whether or not it is necessary, is certainly not sufficient for moral permissibility.

Often, it is not those actually harmed who consent. For example, Shrader-Frechette points out that when the Mescalero Apache accepted a monitored retrieval storage facility for spent nuclear fuel, it was in fact the tribal leaders who agreed and that ‘people who opposed the facility... became victims of retaliation who were likely to lose their housing and their jobs on the reservation’.<sup>53</sup> It is, as has already been stressed, one thing for consent to be given by representatives of those concerned; quite another when those alleged representatives are not democratically elected, and are coercing those they are meant to represent. At the wider level of third world governments, often there is no democratic election process, and/or corruption is rife, rendering it extremely implausible to claim that the individual citizens themselves have consented or even that the consent has been given by their representatives, with their interests at heart.

Even assuming the set of individuals harmed do “consent”, this is rarely free informed consent. Consider the four requirements for such consent, which Shrader-Frechette takes from ‘current... consensus in medical ethics’: risk imposers must disclose full information about the threat; potential victims must be competent to evaluate it; they must understand the danger; and they must voluntarily accept it. Full information is not always disclosed; nor are those affected sufficiently educated to evaluate the risk and understand the danger, if it were.<sup>54</sup>

---

<sup>53</sup> Ibid., 128

<sup>54</sup> Ibid., 77

Acceptance could sometimes be taken as “voluntary”, in the sense that even someone whose only alternative to taking a dangerous job is starvation could be said to be *negatively* free in deciding whether or not to take that job. (They do not have a gun to their head, or someone literally forcing their hand to sign the contract.) However, cases (which no doubt there are) where those accepting exploitative, dangerous employment or hazardous living conditions *are* fully aware of what they are doing *and* consent voluntarily merely serve to bring out the third response.

There is something wrong with a situation in which a person accepts a severe threat to some fundamental interest (such as health) in order to protect some other such interest (for example, nutrition). In such an exploitative situation, an individual human being is having to accept something detrimental to flourishing. The fact that they do, in practice, accept it (they consent) does not protect the person who imposes that harm upon them from negative moral judgement.<sup>55</sup>

I have now addressed three of the potential moral “excuses”. The next section will consider the fourth. However, before leaving this section, I wish briefly to acknowledge one problem facing any argument which hinges on harm to (or consent by) future generations. That is, Parfit’s non-identity problem (NIP).<sup>56</sup> Applied to this case, the problem is as follows. Because a child conceived at one time will have a different identity from a child conceived at another time, future generations if we tackle climate change will not be composed of the same individuals as if we don’t. By maintaining our polluting lifestyles, we cause individuals in future generations to have a life. Given that, unless they have so low a quality of life that it would be better for them not to have lived, how can we have been said to have *harmed* them? Moreover, presumably they would have consented to the action.

Parfit, although he does not provide it, believes that a solution can be found to the NIP.<sup>57</sup> I share his optimism, but it is beyond the scope of this project to

---

<sup>55</sup> Shue makes a similar point, with his example of planting landmines along the Fingerlakes Trail that would explode in several generations’ time: ‘The point is that it does not matter whether some Scouts could be found in 2098 who would agree to walk the Trail for a large enough package of compensation. A society that still acknowledged a right to physical security would not permit such a transaction then, just as we would not permit it now... [V]oluntary agreement is immaterial because such agreements are ruled out by our fundamental attitude to the human body.’ (1999, 41)

<sup>56</sup> Parfit 1984, 358-9

<sup>57</sup> Ibid., 445

offer such a solution.<sup>58</sup> Instead, I will appeal to its far-reaching, and intuitively repugnant, consequences. To take the NIP seriously would render morally permissible almost *any* policy which changes the circumstances of only the as-yet-unborn, so long as it at the same time affected who those future generations were. Similarly, its implications in the consent case indicate, not that this philosophical loophole can be used to justify any environmentally destructive policy so long as its impact is only felt after all those alive today are dead, but rather that consent is not necessarily an appropriate consideration when it comes to actions which affect future persons.

Suppose a malicious scientist were to inject the water supply of a country with a hormone which caused women to miscarry all foetuses save those who were blind. Because the children born (blind) are not the same children who would have been born (not blind) otherwise, it is difficult to say that an individual has been harmed. Moreover, assuming their lives to be worth living, it is at least plausible that the children actually born would consent to this act.

Does this mean that he has done nothing wrong? Surely not. Not only is his action obviously morally wrong but its wrongness can, at least in significant part, be traced to the fact that he has caused a situation to come about in which individual human beings (*those* individual human beings) are limited in their capacity to flourish.<sup>59</sup> He has done something for which the world is a worse place, even if the actual individuals in future generations have not been harmed *in comparison with not living at all*.

Alternatively, consider an individual born in a concentration camp during the Holocaust who nonetheless went on to live a worthwhile life. Is what the Nazis did any less bad because this one individual was not harmed in the narrow sense of being “better off than if Hitler’s policies had not been implemented, because otherwise he would not have been born at all”? All our instincts say otherwise. The wrongness of Hitler’s actions consists in part in the conditions that *this*

---

<sup>58</sup> Weinberg, for example, attempts a solution – one, however, which is specific to a Rawlsian contractualist model. (2002, 406-7)

<sup>59</sup> “In significant part” rather than “entirely” because a part of the wrongness derives directly from his imposing on the women the suffering and health difficulties of miscarriages. There is also the indirect effect on the parents – the psychological pain of seeing their child’s life so limited – but there is a parallel here in the environmental case: new parents, at present, have to deal with the growing awareness that the world in which their children will be adults is likely to be one where capacity to flourish is curtailed by environmental damage.

*individual*, as a small child, had to endure. We object to the conditions imposed by the Nazis to this (as well as countless other) individuals.

Given this, I will skirt the NIP by specifying that when I refer to “harming future generations” I mean: limiting the life prospects of individuals within those generations (whoever they turn out to be) by leaving them with circumstances within which their scope to flourish is limited in comparison with ours.

## VI. Rights and duties

Having rejected three objections to a collective moral duty to address the problem of environmental destruction, there remains Excuse (4), above. This excuse would be used by an *individual* in the face of duties imposed on her by a collective decision-making authority on environmental issues. Such duties, an individual might say, don’t apply to me because they conflict with my rights.

It is not necessary, here, to dwell in any detail on what such “rights” might be supposed to be. I will return to the question in a later section. The point to be made in response to the objection is simply that it is possible to have a right to do something which is nonetheless morally wrong.<sup>60</sup> If I am your landlord and have given you the requisite notice, I am entitled to throw you out of my property, even if it is mid-winter and you and your children will end up freezing on the streets. In such a situation, I have a *legal* right to do something which is, at the very least, morally dubious.

Similar cases can arise even with *moral* rights. For example, if you make me a promise, I have a right to exact the performance of it from you no matter how small a contribution that makes to my interests and how much damage it does to yours. Suppose the promise is to help me move house on a particular day. Were a close relative of yours to fall very ill on that day, I would have a right to insist that you help me, even if this deprives you of a last visit to your relative, and so does you both significant harm.

Thus, it seems perfectly plausible that I might at the same time have a moral right to do something, and a moral duty not to. A moral right is merely a safeguard against others interfering with my performing a certain act; it is *not* a

---

<sup>60</sup> For a discussion of this, see Waldron 1981 or Steiner’s discussion of duty-free zones (1996).

guarantee that, in performing it, I will not be behaving wrongly. Even if we did (although I believe that we do not) have rights which clashed with the individual duties corresponding to a collective endeavour to end environmental harm, it would nonetheless remain the case that we *ought* to fulfil those duties.

## VII. Towards legitimate coercion 1: collectivising the Harm Principle

Considerable ground has already been covered in this chapter. A two-pronged defence (prudential and moral) has been offered of a collective duty to set up those institutions necessary to decide on co-ordinated global action to tackle climate change, which decisions would then be handed down to individuals (and state level actors) as duties binding on them.

However, in *practice*, to have any real chance of success, such an institution is likely to need the power to back up its decisions with coercive authority.<sup>61</sup> A preliminary discussion of how such coercion might be justified will be offered in the following sections, beginning with the morally salient potential collectivity constituted by those collectively responsible, in the weak sense, for environmental harm, then returning to the more difficult cases discussed in Sections I to IV.

The potential coercive enforcement of individual duties correlative to our collective responsibility for environmental harm can be defended using an idea not only appealing and generally accepted, but also acknowledged implicitly by a core tenet of liberalism. The generally accepted idea is that individual freedom *can* sometimes legitimately be limited (that individuals can legitimately be expected to submit to coercively enforced collective authority) to prevent their doing harm to others. The “core tenet” is the Harm Principle, which claims

that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can rightfully be exercised over any member of a civilised community, against his will, is to prevent harm to others.<sup>62</sup>

---

<sup>61</sup> It is likely that this power, in the global institution, would amount to the ability to impose certain sanctions on violating states, leaving the penalising of individuals to those states.

<sup>62</sup> Mill 1859, 14

This principle can be split into two parts:

(Implicit) *The positive claim*: that it is sometimes permissible to limit the liberty of action of an individual human being in order to prevent his harming others.

(Explicit) *The negative claim*: that this is the *only* circumstance in which such interference is permissible.

The focus here is on the positive side. As later sections will demonstrate, it is not my intention to defend the negative claim (which, in any case, is not generally adhered to in so-called liberal societies).

The Harm Principle is standardly couched in individual terms, but can be expanded to the collective level. Here, we can borrow from Andrew Kernohan, who defends an Accumulative Harm Principle:

The state may adopt policies that otherwise would violate neutrality only if individual activity either is, by itself, causing harm to others or is part of an accumulative activity which brings about harm to others.<sup>63</sup>

He defends this principle by responding to three worries, two of which are relevant here.<sup>64</sup> Firstly, that '[t]he individual's own actions may be harmless and would not be liable to regulation if others were not taking similarly individually harmless actions. It appears that the individual is being penalised for what other people do.' This, Kernohan argues, misapplies our intuitions. Individuals are penalised for their own actions, albeit in the context of others doing the same. 'If they had been doing something else instead, no penalty would have been assigned.'<sup>65</sup>

The other concern is that individuals could become subject to 'overly onerous duties'.<sup>66</sup> However, Kernohan points out that similar worries could be raised even with respect to the Individual Harm Principle, without requiring us to reject it altogether.

---

<sup>63</sup> Kernohan 1998, 76. Again, what is relevant here is implicit rather than explicit in the principle: the idea that it is sometimes permissible to violate neutrality when such harm is being done.

<sup>64</sup> Kernohan's other worry focuses on the cases of pornography and cultural oppression, with which he is particularly concerned. It is that an individual will be subject to constraints 'on the basis of his vaguely defined part in a collective activity which brings about harm'. (1998, 83)

<sup>65</sup> Kernohan 1998, 83

<sup>66</sup> Ibid., 83

Suppose a man is wildly flailing about with his fists... [He is] liable to the imposition of a duty which will prevent him from harming bystanders. But what duty? Many duties would succeed in preventing harm: Thou shalt not hit others. Thou shalt not flail about with thy fists. Thou shalt immediately commit suicide. Thou shalt present thyself to the nearest servant of the state for summary execution... Only the first of these, however, is a contender, for though the [Individual Harm Principle] licenses the imposition of duties to prevent harm, it must be supplemented with considerations of justice and efficiency to determine what particular duties to impose.<sup>67</sup>

Similarly with the Accumulative Harm Principle, the point 'is to indicate when it is permissible to impose maximally equitable, minimally onerous harm-preventing obligations'.<sup>68</sup>

Expanding the Harm Principle in this way could legitimate coercive enforcement of collective decisions designed to prevent collective harm. However, unlike Kernohan, who is concerned specifically with the state, I am not restricting discussion to impossible duties within existing political institutions. Rather, I am defending a collective duty to motivate the setting up, if necessary, of an institution with coercive authority. The idea to be borrowed from the Harm Principle is a general one: that individual freedom can sometimes legitimately be curtailed, by *some* legitimate collective authority, when the individuals are part of a set which is collectively harming, in the strong or weak sense.<sup>69</sup>

Hardly anyone would claim that coercion is *always* appropriate in such cases. Rather, the Harm Principle seems to shift the burden of proof, so that it is those harming who have to provide a legitimate reason as to why they should not be coerced out of it.

Thus, within the heart of liberal thought lies a basis for the legitimate coercive enforcement of environmental duties. However, there is an objection to this line of reasoning. Recall the fourth excuse identified at the start of Section V, *Acting within one's rights*. An individual might cite such rights in the face of duties handed down to her by a global environmental decision-making body. The

---

<sup>67</sup> Ibid., 85

<sup>68</sup> Ibid., 85

<sup>69</sup> I use the word "legitimate" very loosely here, intending to include such collective institutions as we have a collective duty to establish, but to make space for the need for those institutions to be established in certain acceptable ways, involving political input from many different states, who in turn are genuinely representing their members.



excuse could then take two forms. Firstly, as already addressed in Section VI, the individual could claim that even if there is an individual duty to act in compliance with the institution, this clashes with individual rights and so is overridden. Secondly, she could claim that any attempt by the institution coercively to enforce the individual duty would clash with her individual right not to be interfered with in certain ways.

As was pointed out in Section VI, an individual's rights set limits on how others may interfere with her. They do not necessarily render her invulnerable to moral claims. However, by exactly that token, the argument of that section is no defence against the second version of the excuse. If I have a right to do something (which is a right against others, to the extent that they are forbidden from interfering with me), then surely I cannot be coerced into not doing it even if it causes harm.

This has long been acknowledged by proponents of the Harm Principle. As already stressed, Mill's principle does not render the causing of harm to others a sufficient condition for legitimate interference. Rather, he argues: 'As soon as any part of a person's conduct affects prejudicially the interests of others, society has jurisdiction over it, and *the question of whether the general welfare will or will not be promoted by interfering with it, becomes open to discussion.*'<sup>70</sup> Corresponding to this "general welfare" is a system of *social* rights,<sup>71</sup> and it is only when harm to others falls outside the sphere of what these give the harming individual a right to do, that the state can interfere with individual freedom.<sup>72</sup>

This thesis, however, does not offer a purely utilitarian argument. Thus, rather than remain within Mill's framework of *social* rights grounded in overall welfare, the question here is more general: what rights might individuals in this generation have such that, while it is unfortunate that what they are doing results

---

<sup>70</sup> Mill 1859, 83-4 [emphasis added]

<sup>71</sup> That is, those established by the constitution of a political society and designed to bring about the greatest good to the greatest number. Mill eschews so-called natural or human rights: 'I forgo any advantage which could be derived to my argument from the idea of an abstract right, as a thing independent of utility.' (1859, 15; emphasis added)

<sup>72</sup> For example: 'Whoever succeeds in an overcrowded profession, or in a competitive examination; whoever is preferred to another in any contest for an object which both desire, reaps benefit from the loss of others, from their wasted exertion and their disappointment. But it is, by common admission, better for the general interest of mankind, that persons should pursue their objects undeterred by this sort of consequence.' (Mill 1859, 105)

in harm to others, they have a perfect right to do it, the others have no right that they not do it, and they cannot legitimately be coerced into not doing it?

An unrestricted right to negative freedom is clearly implausible. Not only would it undermine the whole idea of the Harm Principle, but it would not be able to carry the argument. There are negative freedoms at stake on both sides of the debate. The negative liberty of the woman, in Cohen's example, who wants a Selfridges sweater but lacks the money to pay for it and so will be prevented from walking out of the shop with it, is as violated as that of the rich American who is prevented from putting leaded petrol into his car unless he pays tax on it.<sup>73</sup> The liberty of a resident of a Mexico City shanty town is restricted if he attempts to take his asthmatic child into a private health spa, but is forcibly stopped.

Nor will self-ownership, alone, do the necessary work. For this generation to have an all-trumping right to continue to pollute, what are required are not only strong self-ownership rights, but also some very stringent ownership rights over natural resources – rights strong enough to justify not only the using up of such resources by current generations, but also the devastating impact this use has on potentially unlimited resources such as water and air. It is unclear that there could be a plausible derivation of such rights and, given the severity of the harm in question, surely the onus is upon opponents of this line of argument to come up with such an account.

In its absence, and because full discussion of this point would require a detailed analysis of property rights, which is beyond the scope of this project, I will simply acknowledge it as a key assumption that we do not have any property rights over natural resources which are strong enough to excuse collective environmental harm – or absolve us from the individual responsibilities which come with it. This leaves us with one significant argument the less against the legitimacy of coercive enforcement of those duties.<sup>74</sup>

---

<sup>73</sup> Cohen 1995, 58

<sup>74</sup> Again, it is worth stressing that I do not claim this removal of one significant counterargument as a complete defence of the legitimate coercive authority of such institutions.

## VIII. Towards legitimate coercion 2: the Principle of Fairness revisited

Coercive enforcement of our negatively-grounded duties (subject, of course, to certain conditions about *who* does the coercing) is relatively uncontroversial, accommodated as it is within the central tenants of liberalism. Not so coercive enforcement of our prudentially-grounded incentives, or our positively-grounded moral responsibilities. Chapter Four indicated that enforcement of collective decision making could be as necessary as the collective decision making itself for the achievement of the common interests, goals or purpose holding many collectivities together. But this in itself does not *justify* collective coercion.

Hardin, discussing the population problem as a game theoretic case (the Tragedy of the Commons) recommends ‘mutual coercion, *mutually agreed upon*’.<sup>75</sup> Building on Section II, a similar solution might be proposed in the general environmental case: we all have a prudential, rational incentive to consent to a coercive institution to address climate change (including overpopulation). This is because we are in a situation analogous to Hobbes’ state of nature. We all know it is rational to act to address climate change as part of a collective effort to do so, but we all also know that there is an individual incentive to cheat on an agreement to act collectively, unless that agreement can be coercively enforced. Thus the “necessary collective decision-making apparatus”, to which I referred in Section II, would in practice be one with coercive authority to enforce its decisions.

The Hardinian solution, then, is that we establish such an institution. Its coercive authority would be legitimised by precisely that universal informed consent (an uncontroversial contractualist line of reasoning). Thus, we turn our prudential incentive into obligations which that institution can legitimately coercively enforce.

Leaving aside the question, familiar to Hobbes theorists, of *how* the step to acceptance of a particular coercive institution is supposed to take place, even if all *are* motivated to consent to some form of coercive authority,<sup>76</sup> there are significant difficulties with applying this model to the actual global

---

<sup>75</sup> Hardin 1968, 1247 [emphasis added] Similarly, recall that Runciman & Sen call for an ‘enforceable contract’ (1965, 556) between the parties in their analysis of the Prisoner’s Dilemma as giving a substantive content to the idea of the “general will”.

<sup>76</sup> I return to this in Chapter Six, Footnote 38

environmental predicament: not least, the many who would not consent to any such scheme.

Firstly, some members of the collectivity would deny their ATC interest in a collective effort to address climate change and withhold their consent. These were the kind of cases discussed in Chapter Four, Section III, under the Ignorance Argument, and recalled in Sections II and III of this chapter with the example of Young George. Secondly, as Sections I and IV made clear, there are some members of current generations who genuinely *aren't* part of the collectivity with respect to the goal of environmental preservation, as it is neither their goal nor in their ATC interest.

Turning to the small print, Hardin, no doubt well aware of such cases, does not actually require unanimity. Rather, he proposes coercion 'mutually agreed upon by a majority of the people affected'.<sup>77</sup> This, however, is a much more controversial proposal, and lacks the justificatory force of informed consent by all. Essentially, a coercively-enforced collective decision-making institution would be imposed on these individuals who did not consent. Can this ever be justified?

Consider the two problem cases again, starting with Young George. Again, we might appeal to the Modified Principle of Fairness. Central capabilities are at stake, rendering the good a presumptively beneficial one as specified by Klosko's first condition. Moreover, his second condition (that the benefits of co-operation outweigh the burdens) is satisfied not only in his weak sense (for the community overall) but in the strong sense (for that individual).

However, we should be wary of too quick a step from the defence of a moral obligation to the claim that coercive enforcement of that obligation is legitimate. According to Hart's original paper, legitimate enforceability of the duty to co-operate in a scheme from which one benefits goes hand in hand with the duty itself. Klosko similarly slides the two together. But Nozick has objected that the right of enforceability would not automatically follow from the obligation to contribute, even if there were such an obligation (which he, as we have seen, denies).<sup>78</sup>

---

<sup>77</sup> Hardin 1968, 1247

<sup>78</sup> Nozick 1974, 91-2

Certainly, it is not always obvious that an obligation should carry a right of enforceability. Consider the special obligations that we acquire on a day to day basis. If I promise to post your letter for you, I have an obligation (to you) to do so. However, if you happen to notice me about to pass the letter box without posting it, are you entitled to grab hold of me and force me to put your letter in the box, or imprison me if I fail to do so? Is society entitled to do so? It seems not.

The question, then, is whether a special case can be made for the enforceability of obligations incurred under the Principle of Fairness – or at least under the Modified Principle discussed in Section III.

The dilemma is as follows. Prudence and fairness both dictate that Young George should co-operate. However, he is motivated to refuse to co-operate, not simply because he wants to free ride on others' efforts but because he believes he would be better off without the overall scheme than co-operating within it. On what grounds can the rest of the collectivity legitimately coerce him?

They might do so by appeal to his own fundamental interests. Environmental damage constitutes such a serious threat to his being able to live a full, flourishing life at a later stage that perhaps he can legitimately be coerced into contributing to a collective effort to tackle climate change.

There is an obvious objection to this line of argument: it seems paternalistic. If all we can say to Young George to justify interfering in his choice of lifestyle is that it will be better for *him* although he doesn't know it, we are on controversial ground. Certainly, we cannot appeal to the forefather of liberalism to help us, as we could in the previous section. Mill's own example makes this clear.

If either a public officer or any one else saw a person attempting to cross a bridge which had been ascertained to be unsafe, and there were no time to warn him of his danger, they might seize him and turn him back, without any real infringement of his liberty; for liberty consists in doing what one desires, and he does not desire to fall into the river. Nevertheless, when there is not a certainty, but only a danger of mischief, no one but the person himself can judge of the sufficiency of the motive which may prompt him to incur the risk; in this case, therefore (unless he is a child, or delirious, or in some state of excitement or absorption incompatible with the full use of the reflecting faculty), he ought,

I conceive, to be only warned of the danger; not forcibly prevented from exposing himself to it.<sup>79</sup>

The message is clear: we warn the man if we can, and let him choose his own level of risk exposure: forcible intervention is only legitimate when it isn't possible to warn him in time.

This thesis has already tackled the paternalism objection: Chapter Three defended the capabilities approach, as an attempt to get at what is involved in a fully flourishing human life, against the charge of unacceptable paternalism. It did so on the twofold grounds that the thick conception of the good inherent in something like Nussbaum's list was a vague one, and that capacity to develop and live by one's *own* conception of the good and rational plan of life was central to the approach, rendering it more likely to leave individuals able to make long- and short-term autonomous choices than a "thin" liberal account which focused exclusively on negative freedom.

Such considerations go some way towards softening the accusation of paternalism here. We might question whether Young George is really making an autonomous decision, against a background of all the capabilities. Perhaps he has been brought up to be dependent on a certain kind of lifestyle. Equally, one might argue that his choice does not constitute a rational *decision* at all: he is not so much weighing all the options and choosing between them, as flatly refusing to acknowledge the reality of the implications. (As with the Head-in-the-Sand Farmers.<sup>80</sup>)

In either case, the arguments of Section V, Chapter Three, constitute a strong case against the claim that it would be unacceptably paternalistic to maintain a political structure which prevents his acting on his decision.

However, if we could not criticise his decision on either of the two grounds above, Young George could constitute a particularly tricky case for the capabilities approach. Suppose he has grown up with all the capabilities. As part of this, he has had opportunities to enjoy the natural world and access to non-

---

<sup>79</sup> Mill 1859, 107. Interestingly, we could appeal to Mill for help in certain cases: he condones the overruling of an individual's decisions to deprive himself with certainty of future liberty – i.e. to sell himself as a slave – on the grounds that such a decision defeats the very purpose for which he is allowed authority in all cases relating only to himself (the preservation of liberty). (1859, 113-4). At least arguably, the same reasoning could be used to veto suicide.

<sup>80</sup> Chapter Four, Section III

high-energy pursuits. Moreover, he does accept the serious risk to his own central capabilities posed by climate change. However, he decides that living a certain lifestyle in his youth is worth that sacrifice. This recalls the problem case addressed in Section VI of Chapter Three: an individual who had hitherto had all the capabilities, and so satisfied the Razian conditions for autonomous choice, but decided permanently to give one of them up.

(I would, however, reiterate the point made then as to the rareness of such cases. It is far more likely that he simply refuses to acknowledge the reality of his decision – he ignores the facts, or, as Papineau suggests, he needs to reconcile his desires and factual information and it is easier for him on this occasion to adjust his factual beliefs.<sup>81</sup>)

In Chapter Three, I identified two versions of the capabilities approach, both plausible: the Permissive Version, according to which such a person would be said still to have a fully flourishing life; and the Non-Permissive Version, according to which it would remain in his ATC interest *not* to expose the central capability to serious risk. In this latter case, I appealed to the vague, open-ended nature of the central capabilities (and their being capabilities rather than functionings) to save the account from unacceptable paternalism.

Now, however, it becomes apparent that this distinction has serious implications. On the Permissive Version, the capabilities approach would no longer give us grounds for saying it was in Young George's ATC interest to co-operate. He would then count as a George-like case, not actually in the collectivity. On the Non-Permissive version, while it would not be unacceptably paternalistic to *criticise* his choice, given that he made it autonomously, it is harder to defend *forcing* him to retain the capability, at cost to himself, purely on the grounds that it *is* in his interest to do so. There is a difference between judging an autonomous person's quality of life according to a conception of the good which they do not share, and directly coercively imposing that conception on them, in the sense of restricting their freedom in certain ways in order to ensure that they have opportunities which they do not want.

---

<sup>81</sup> Papineau 1978, 168. See also Chapter Four, Section III

At the very least, examination of this case indicates that something more than Young George's own (denied) ATC interest might be needed to justify coercing him into participating in the collective effort.

Now consider George. Given that we cannot demonstrate that it is in his ATC interest to co-operate, we need more than a paternalistic case for coercing him into doing so. Section IV suggested that he had a positive moral obligation to submit, grounded in the duty to prevent harm to the others. This also supplemented the case for Young George's obligation.

My claim is as follows: once again, what makes the difference between this case and examples such as Nozick's public address system is the fundamental nature of the interests at stake – as interests of the others involved. It is the harm that would be prevented by co-operating which lends plausibility to an enforceable Principle of Fairness on the Modified and even the Weak Modified Version.

To make this point, Nozick's own argument can be turned on its head. According to Nozick, it is especially repellent to claim that others could enforce my supposed duty to contribute to something which benefits me, if those others benefit as well.<sup>82</sup> This often has an intuitive appeal. For example, we might find it morally questionable that I force my friend to go to the gym, for the benefit of her health. If the gym is mine and she incurs a heavy membership charge, we find it even more so.

However, our intuitions are reversed if the "benefit" to those others is not something they could easily do without, but prevention of some grave harm. Consider the following variant on Mill's bridge case: Mill and I are both on a rope bridge when it collapses, leaving us hanging. We could, by careful concerted effort, climb up to safety. However, Mill, a risk-lover, prefers to run the (much higher) risk of dropping into the canyon below. If he doesn't move, I can't, so I will have to take the risk as well. It is at least arguable that I would be entitled to force him up against his will simply on the grounds of saving his own life, but it is much more obviously (intuitively) legitimate for me to oblige him to co-operate if otherwise it is extremely likely that we would *both* die. Even if the

---

<sup>82</sup> 'On the face of it, enforcing the [Principle of Fairness] is objectionable. You may not decide to give me something, for example a book, and then grab money from me to pay for it, even if I have nothing better to spend the money on. You have, if anything, even less reason to demand payment if your activity that gives me the book also benefits you.' (Nozick 1974, 95)



situation is changed so that the risk to him is less (we would fall into a river and he can swim fairly well, but I can't) and the costs of co-operating are higher, I am inclined to think that I am entitled to force him to co-operate (depending, of course, on just how high the costs of co-operation are).

Similarly, given the cost to the rest of us if the Young Georges of the world do not contribute to a collective effort to tackle climate change, and given that it is in their own ATC interest that the effort take place, there is a moral case to be made, in line with the Modified Principle of Fairness, for our coercing them into doing so. With George, the moral case must rest more heavily on the costs to us if he does not co-operate, but the fact that there is a benefit to him, in rendering his fundamental interests more secure, does at least add some weight to the argument.

This is very much an outline argument, and does not pretend to be as strong even as that of the previous section, which is in itself only a first step towards the justification of some specific coercive collective environmental institutions. Positive duties are controversial enough; rights of enforceability even more so. However, the cases in which I have attempted to fill out the intuitive appeal of an enforceable Principle of Fairness draw attention to the fine line between harming and not helping. Our intuition, insofar as it does legitimise coercion in such cases, does so by bringing them close to those covered by the Harm Principle. Moreover, there is also *some* gain, in such cases, to the relevant individual. All I have done here is suggest, somewhat tentatively, how these points might come together to defend the coercive enforcement of the Modified and even the Weak Modified Principle of Fairness.

## **Conclusion**

This chapter has defended the collectivityhood of most of humanity with respect to an interest in environmental preservation, and used this to give a prudential incentive for the promotion of (and compliance with) a collective decision-making body. This was supplemented with a moral incentive grounded in the Principle of Fairness.

However, it has acknowledged that some members of the global elite, above a certain age, may genuinely not be part of this collectivity. A case was made,

appealing to positive moral duties, for their duty to co-operate. More strongly, a negatively-grounded moral obligation to address climate change was offered: many of us (including those who can exclude themselves from the actual collectivity), through our contribution to environmental damage, constitute a morally salient potential collectivity.

This claim was defended against various moral “excuses”, which might be used to claim that we are nonetheless acting morally permissibly. In the course of this, a step was made towards defending the legitimacy of coercive enforcement of decisions by a global environmental decision-making body, and one important counterargument was rejected. The question of legitimate coercion of the prudential or positively grounded duties was also considered, and an argument was outlined by which a modified Principle of Fairness might be supplemented to justify coercion in certain cases, including the environmental one.

The next chapter will complete the application of the model developed in the first half of this project, to the environmental case. It will outline the institutional implications of my account, and address some of the most likely objections.

## Chapter Six

### Implications and Objections

Hitherto, I have defended a model of a collectivity which is broader than the conventional, intentionalist account of a social group. A set of individuals constitute a collectivity if they are mutually dependent through some shared or common goal or ATC interest, in the sense that that goal or interest requires shared or common action for its achievement.

I have also defended the notion of morally salient potential collectivities. These meet two conditions. Firstly, they are sets of individuals who, were they to adopt a certain goal, would become a collectivity. (The goal is a common one in the sense above.) Secondly, they ought (morally) to adopt the goal. Filling out this *ought*, I have identified at least one way in which a potential collectivity can be morally required to adopt the goal: if its members are collectively responsible (including in a weak sense of collective responsibility, through the predictable aggregation of individual acts) for unacceptable harm to others; and if the goal is the prevention or mitigation of that harm.

Political philosophy, I have suggested, should be concerned not only with the wider set of collectivities (rather than the narrower one of social groups), but also with morally salient potential collectivities. For the purposes of assignment of duties, they should be treated as actual collectivities, because they have a collective moral responsibility to adopt the goal that would make them so. Moreover, I have argued that, for many collectivities, it is “better” (better for achieving the shared goal or interest) that the collectivity itself be the unit of agency for decisions relating to that goal or interest, than that it be left to the combination of individual decisions and actions.

The previous chapter defended the claim that the whole of this generation of humanity is included in one or other (usually both) of an actual collectivity and a morally salient potential collectivity, by virtue of environmental concerns. It then made a prudential case for the promotion of and adherence to collective decision-making institutions, and supplemented this with a moral argument (based in part on considerations of fairness, backed up by positive moral duties, and in part on negative moral duties). An outline was offered of an argument for the coercive enforcement of environmental duties.

This chapter brings together all these strands. Its basic suggestion, elaborated in Section I as the Decision-Making Claim, is that each collectivity in which humans find themselves should make precisely those decisions *qua* collectivity which both: are directly related to achieving or satisfying the common purposes, interests or goals of the collectivity, *and* need to be made collectively to achieve better those goals, purposes or interests. I am advocating neither a global government which takes over all the current responsibilities of the state, nor that those states (as collectivities) should dictate to citizens in all aspects of their private lives. However, I do not claim to offer a conclusive argument for this limitation. My primary objective is to defend the conclusion that state sovereignty should be weakened, in that control over certain fields should be ceded to a global decision-making body.

Accordingly, Section II will borrow from Simon Caney, in the global justice debate, to show how the institutional framework defended here would weaken state supremacy. The rest of the chapter will address various potential objections, to this framework and to the overall thesis. A full rebuttal of some of these lies beyond the scope of the project, and I can only point the way for future discussion. I will respond to others in more detail. The objections are as follows.

*Objections to the Decision-Making Claim:*

- Communitarian objection (Section III)
- Cosmopolitan objection (Section III)
- Statist objections (Sections III to V):
  - Liberty-based argument (Section III)
  - Society of states argument (Section III)
  - Same benefits argument (Section IV)
  - Moral burdens argument (Section V)

*General objections:*

- Moral burdens argument (Section V)
- Past polluters objection (to the collective harm defence of environmental duties) (Section VI)
- Too inclusive objection (Section VII)

## I. The Decision-Making Claim

Each of us, as an individual human being, belongs to many different collectivities – voluntary, involuntary, intentional, unintentional, small, large, and so on. Some of these collectivities intersect; some are subsets of others. We are members of families, friendship groups, local communities, businesses, academic institutions, nations, states (often nation states), federal groupings of states such as the EU, global economic networks and, crucially, the whole of humanity.

The space left for freedom of decision making at each sub level will be affected by decisions made at the level of the bigger collectivities of which they are a part. This raises the question of *what*, exactly, should be decided at the global level, thereby limiting the potential courses of action available to individual states. Further down the line, to what extent should states be able to dictate to their citizens? This chapter will apply the arguments of previous chapters to these questions. As with Chapter Five's appeal to the Principle of Fairness and the Harm Principle, I will draw on a key liberal thinker (in this case Dworkin) to support my case.

Chapter Four identified six scenarios in which it would be better in terms of satisfying the goals, interests or purposes through which a collectivity is made up, to make certain decisions with the collectivity, rather than individual members, as the unit of agency – i.e. to make those decisions collectively and hand them down as restrictions on individual action. This prompts the following claim:

### *Decision-Making Claim*

Each collectivity in which humans find (or put) themselves should make, *qua* collectivity, precisely those decisions which:

- are directly related to achieving or satisfying the common purposes, interests or goals of the collectivity; *and which*
- need to be made collectively in order better to achieve those goals, purposes or interests.

These decisions should then be handed down as restrictions on individual (or subset collectivity) behaviour.

Note, however, the following points. Firstly, the claim does not mandate collective decision making in cases where, although individuals depend on each other for achievement of some end, that end will in fact come about automatically from the individual pursuit of individual ends. Convergence Cases in game theory provide one example.<sup>1</sup> Alternatively, consider a number of people with the shared goal of preserving the local bat population, when in practice the best way to bring this about is for each household, separately, to encourage the nesting of bats. However, the range of arguments in Chapter Four suggests that such cases are likely to be relatively rare, especially as collectivities increase in size, while Chapter Five demonstrated that the environmental case does not fall into such a category.

Secondly, this claim is two-sided and only one side (what might be called the positive side) has been sufficiently defended in this thesis. I have argued that it is better in terms of the goals, purposes or interests through which the individuals are bound together into the collectivity if *at least* these decisions be made qua collectivity, whether formally or informally. A full defence of the other side of the claim – that *only* such decisions should be made at such a level – is beyond the scope of this project, although I will touch on some of the theoretical ideas on which it draws.

The Decision-Making Claim is not a new idea, far from it. It is an expansion of an idea defended by Dworkin, albeit one which goes well beyond what Dworkin himself would allow. This idea is that the areas on which a community should dictate to its members should be limited to those within the ‘communal

---

<sup>1</sup> ‘The interests of the two players converge but they do not coincide. They do not coincide since each player harms the other by pursuing his own interest. Their interests converge, however, in that if each picks his dominant strategy, they both do better than if both did not. The harm each can do is small enough that it makes sense to allow it to occur. Altruism here would be misguided.’ (Hamburger 1979, 82-3) Note, further, that in some, generally small, *ad hoc* collectivities, the required co-operative action takes place without any conscious collective decision making. Recall Sheehy’s escaping prisoners (Chapter Two, Section I), adapted so that each believes himself to be rowing alone. It is by chance that the prisoners are rowing in time and succeed in achieving the group goal, and they do so without being aware that theirs is a collective endeavour. Alternatively co-operation might be so automatic as barely to register at a conscious level. Recall Hume’s rowers (1777, 306-7) or the campers taking down their tent in Chapter Four, Section X. Such cases would not require acknowledged or formal collective decision making unless the almost unthinking nature of the co-operation meant that an individual could easily be distracted, undermining the collective effort. Consider Rousseau’s example of savages hunting a stag: each individual would be inclined to run off at the sight of a hare. (1755, 86-7) The way to motivate him would be by appeal to his long-term interest in being fed reliably, which is best met by submitting himself to the collective unit of agency in hunting.

life' of that community, the success or failure of which 'is part of what determines whether its members' lives are good or bad'<sup>2</sup>.

While I agree with Dworkin in rejecting the *metaphysical* claim 'that communities are fundamental entities in the universe and that individual human beings are only abstractions or illusions',<sup>3</sup> this thesis has defended a broader idea of what can count as the common, or collective, "life" of a community than Dworkin's own practice view. For Dworkin, for a community to have a communal life, not only must that community's composition centre around some collective act or acts, but individuals' attitudes must reflect this: individual acts are 'performed *self-consciously*, as contributing to the collective act'<sup>4</sup> This thesis has rejected such a limitation, defending a broader, non-intentional understanding of the "common life" of some group of individuals as integral to how well individuals' lives go.

The Decision-Making Claim can also be seen as a variant on the principle of subsidiarity, familiar from early debate on the European Union. This 'regulates authority within a political order, directing that powers or tasks should rest with the lower-level sub-units of that order unless allocating them to a higher-level central unit would ensure higher comparative efficiency or effectiveness in achieving them'.<sup>5</sup>

According to this account, decisions relating to the protection of the environment should be made globally, through the establishment and maintenance of some global environmental decision-making body. (So, relatedly, should some decisions regarding international security, for example concerning the development of nuclear power.)

Legitimate states would make the decisions related to the common interests or goals which render them collectivities, subject to the constraints laid down by that body. These would be so-called "political" decisions, by which I mean, with Dworkin, executive, judicial and legislative decisions relating to national distributive justice and the corresponding tax regime, to the criminal justice system, and to national security.<sup>6</sup> With federal groups such as the EU, decisions

---

<sup>2</sup> Dworkin 2000, 223

<sup>3</sup> Ibid., 225

<sup>4</sup> Ibid., 227 [emphasis added]

<sup>5</sup> Føllesdal 1998, 190

<sup>6</sup> Dworkin 2000, 228

relating to the common interests or goals of members of that collectivity would be made at the collective level, with the individual states' policy making constrained again by any such decisions.

Sub-state communities, such as religious groups, would make decisions relating to their common interests or goals but, again, within the boundaries set by the wider-level decisions. The state would not dictate a person's sex life, or whether she observed a certain religion, or whether she ate Frosties or Coco Pops for breakfast, although membership of some community within the state might come with further restrictions on action. Being either a Muslim or a member of The Vegetarian Society, for example, the individual might be restricted in her choice of breakfast to the extent of not eating bacon.

Sub-state collectivities, on the structure I envisage, are restricted in enforcing their collective decisions. Such collectivities can, collectively, impose sanctions (ultimately exclusion from the collectivity) but they cannot legitimately interfere with an individual's freedom, property, or person, or back up their penalties with such a threat. Importantly, individuals have the option of leaving such sub-state collectivities. If Stella is a member of the Vegetarian Society and she decides to start eating fish, she may only be able to do so on condition of quitting the society, but it (the collectivity) cannot force her to stay and comply with the no-fish requirements.

This gives the state (or arguably some federal organisations) the role of protecting members of sub-state collectivities against attempts, by those collectivities, coercively to enforce collective decisions, rather than allow members to opt out. But this role – of protecting, ultimately, the security of the person – already falls into the realm of the common interests (security, peace, etc) which bind us as members of the state. Moreover, precisely because it *is* in such involuntary collectivities as the state, where there is generally no opt-out clause, that coercive authority to back up collective decisions can be necessary, there is a strong incentive not to expand the state's sphere of influence any further than is absolutely necessary. That is, for fear of imposing some precise, comprehensive conception of the good on individuals in a collectivity from which they *cannot* opt out, political philosophers would be wise to err on the side of caution.



Having thus expanded on the Decision-Making Claim, the next section will indicate the kind of institutional change it is likely to require.

## II. Weakening state supremacy

I draw in this section on Caney's proposal for global institutions to administer distributive justice. I do so both because of the clear framework he provides for debate, and because the institutional structure envisaged on my account has considerable similarities to that he advocates.

Caney lays out four defining features of state sovereignty: *legality*, or authority to coerce, *supremacy*, or absolute authority (the absence of final authority elsewhere), *territoriality*, or a territorially defined group of people over whom its supreme authority is exercised, and *comprehensiveness*, or ultimate authority over all policy areas.<sup>7</sup>

The model defended here, like his, denies states comprehensiveness (and, accordingly, denies them supremacy in the sense of being the final authority for all the decisions within their remit). However, the powers assigned to global institutions differ: I would allocate them (or it) power over environmental decision making without, for the purposes of this work at least, drawing any conclusions regarding distributive justice.<sup>8</sup> Caney is concerned to establish international institutions to apply principles of international distributive justice.

Moreover, Caney, defending the weakening of comprehensiveness on cosmopolitan grounds, appeals to both rights-based and instrumental considerations.

Take... the claim that people have a democratic right to exercise control over the factors that affect their rights. Some of these factors are likely to be global in nature... whereas other factors (say, people in their vicinity impacting on their right not to be assaulted) are more local. Given this, there is a case for different institutions operating at different levels and addressing different policy issues... Whereas, some matters like cultural issues are best dealt with by local – i.e. “sub-state” – authorities, others which cover a wider range of

---

<sup>7</sup> Caney 2007, 149-50

<sup>8</sup> The exception would be redistribution as called for as a result of past environmental harms (duties of reparation)

issues (like protecting the environment or administering cosmopolitan principles of fair trade) may be more effectively dealt with by “supra-state” authorities...<sup>9</sup>

He adds, of supremacy, that ‘since it is in practice impossible to keep policy areas utterly separate, a system in which no institution has comprehensiveness also lacks supremacy. There may be no one political institution that has final authority.’<sup>10</sup>

Quite apart from the fact that I would count as global such factors as environmental protection, rather than (necessarily) ‘the global economy impacting on [individuals’] right to subsistence’,<sup>11</sup> there are some important distinctions to be made here. Firstly, my argument is not “rights-based”. Like Judith Lichtenberg, I suggest that appeal to rights is ‘empty’<sup>12</sup> without the grounding of those rights in precisely the sort of considerations discussed above – that is, of fundamental interests according to some kind of a model of human flourishing.<sup>13</sup> Given that, I prefer to refer directly to those interests. However, the capabilities approach *does* incorporate as a functional capability control over one’s environment, including one’s political environment.<sup>14</sup>

Secondly, I would quibble with the suggestion that there is no ultimate authority. Rather, ultimate authority lies in different institutions for different decisions. Global institutions would be the supreme authority on environmental decision-making, in that they would establish targets and so forth, and have the authority to enforce their decision through various sanctions on noncompliers. This leaves a sphere of influence for individual states, regarding political decisions, which in turns leaves much to individuals and sub-state collectivities.

It remains an open question at this stage exactly how much detail will have to be gone into at the global level, and how much of the detail of implementation can be left to sub-global collectivities. However, it would be a strong claim – and one requiring more argument than offered in the quote above – that, because the state is subject to the authority of global institutions on environmental issues,

---

<sup>9</sup> Ibid., 163

<sup>10</sup> Ibid., 163

<sup>11</sup> Ibid., 163

<sup>12</sup> Lichtenberg 1981, 82

<sup>13</sup> In fact, Caney himself also favours broadly “well-being based” justifications of civil and political human rights’. (2007, 72)

<sup>14</sup> Nussbaum 2000b, 80

there are *no* issues on which it is the court of last appeal. If supremacy is to be understood as a separate consideration from comprehensiveness (and originally this is how Caney lays out his model) it is surely at least *prima facie* possible that different political institutions can have supremacy in different policy areas.

### **III. Objections to the Decision-Making Claim**

Having completed the model upheld by this thesis, by drawing out its implications in terms of the kind of institutional framework we are understood to have a duty to bring about, I will end by acknowledging various objections. I will attempt to answer some here; others have already been answered, at least in part; others still I raise only to acknowledge that a full response is beyond the scope of the current project. However, I will at least suggest how a response might be framed. This section begins with objections to the Decision-Making Claim and the corresponding curtailment of state authority.

#### *The communitarian objection*

It might be argued that, rather than be limited to the common interests or goals of the collectivity in a relatively narrow, political, sense, the state should be able to impose some ethical “code” on individuals, for their own good, or for the good of others in that state community. Thus homosexuality or abortion might be banned. Others would argue that, while state powers should be limited to the “common life” of the state, this common life is much broader than a restriction of state authority to the realm of the “political” would imply. This is what Dworkin dubs the illiberal argument from integration.<sup>15</sup>

As already stressed, I do not claim to offer a comprehensive defence of both sides of the Decision-Making Claim. It is beyond the scope of this project to respond in detail to the communitarian arguments. However, I would note that

---

<sup>15</sup> Dworkin 2000, 222-30. Dworkin criticises four “illiberal” arguments for the expansion of state authority to cover sexual orientation: an argument (based on the harm principle) from democratic theory; an argument (based on promoting well-being) from paternalism; an argument from self-interest, which suggests that liberal tolerance renders political communities unable to meet members’ material or intellectual needs; and the argument from integration, his response to which was touched on in Section II. See Dworkin 2000, 212-230. The article, “Liberal Community”, originally appeared in the *California Law Review* 77 (1989) and was a response to the now infamous decision in *Bowers v. Hardwick* (1986), in which the US Supreme Court upheld the constitutionality of a Georgia sodomy law.

the limitation of authority at each level follows naturally from the model built up in previous chapters: one on which *individuals* are the units of moral concern. Collective authority has been advocated *only* in situations where: this is the best way to achieve the goals (selfish or otherwise) through which all members of the collectivity are mutually dependent, it is necessary for the protection of some fundamental interest (central capability) of all those individuals, *or* there is collective responsibility for serious and morally unacceptable harm to others.

The onus would thus be on the communitarian either to appeal to one of these arguments to justify (say) a ban on homosexuality,<sup>16</sup> or to defend more extensive powers of collective decision making, without threatening the centrality of the individual, flourishing life (which flourishing, recall, involves a core element of autonomy, or planning one's own life against a background of reasonable options).

### *The cosmopolitan criticism*

Just as communitarians and (we shall see below) statist would attack me for curtailing state sovereignty, those in the cosmopolitan camp in the global justice debate might attack me for not limiting it *enough*. Not only, they would say, should environmental (and international security) decisions be made globally, but so should other decisions, notably those relating to distributive justice. Some might even go further, to defend a global state to replace the current system of nation states.

I have deliberately left open the question of whether duties of distributive justice extend beyond the nation state. It is possible that global duties of justice could be defended in the same way as globally assigned environmental duties.<sup>17</sup> One point in favour of such an extension is that, as briefly discussed in Chapter Five, Section VII, there is little reason to believe that western states, corporations or individuals, *or* elites in other countries, have anything like the strong and

---

<sup>16</sup> This seems unlikely. Firstly, it is certainly no part of the capabilities approach that it is in an individual's fundamental interest *not* to be homosexual. Quite the reverse: freedom of choice in such matters is central. Secondly, preventing homosexual activity is not a genuinely shared goal: it is not a goal to which all adhere. Thirdly, *even* if (which I would query) living in a state in which there are also homosexuals were to constitute harm to some religious types, there is no question that the harm done to the homosexuals by repressing them would outweigh this (recall the moral excuses of Chapter Five, Section V) and undermine any attempt to ground interference in collective harm arguments.

<sup>17</sup> I will come back to this in Section VII.

extensive property rights in natural resources which they currently claim. However, that need not mean that global duties of distributive justice would be as extensive as those within each state (although they could plausibly – and would on a parallel account to that advocated above – take priority over such duties). More generally, a very strong case would have to be made before all, or even many, of the powers currently exercised by state governments would have to be ceded to a global equivalent.

### *Statist objections*

Like Caney, I am advocating a system in which states lack both supremacy and comprehensiveness.<sup>18</sup> This raises objections from the statist camp. This, for the purposes of this project, can be taken as designating those who condemn any weakening of state sovereignty. These are the same type of arguments made against a cosmopolitan model which includes international institutions for the application of international distributive justice. Following lines taken by Caney and Darrel Moellendorf, there seem to be four key objections:<sup>19</sup>

- *Same benefits argument*: there is no need to make decisions at the wider collective level, as the optimal result will be achieved by the actions and co-operation of states.
- *Moral burdens argument*: this model, in obliging individuals to consider interests geographically or temporally distant, places too heavy a burden on individuals.
- *Liberty-based argument*: global political institutions, in centralising power, are detrimental to individual freedom.
- *Society of states argument*: for individual human beings to flourish, it is necessary that they be members of a state within a “society of states”.

The first two arguments will be debated in the next two sections. With regard to the last two, the *liberty-based* argument, as Caney points out, is indiscriminating: it seems to allow for no possible global political institutions but a world state, against which it would have most force. Moreover, as he also

---

<sup>18</sup> Caney 2007, 163

<sup>19</sup> Ibid., 148-82. Moellendorf 2002, 30-67

notes, the current system frequently oppresses individuals to the detriment of their freedom.

To this we might add that, according to the view defended in Chapter Three, what is *valuable* is not negative freedom *per se* but its contribution to individual flourishing, in particular insofar as it is a component of an autonomous life. The collective decision making advocated on my model restricts individual freedom of action only when that action, understood as contributing to a collective act, would contribute to the undermining of individual flourishing, either for the whole collectivity, including that individual, *or* for some other person(s) to such a level as to constitute morally unacceptable harm.

With regard to the latter, note that libertarians themselves have been remarkably inadequate on the question of restricting harm, and have had even less to say regarding the serious harms which result from the aggregation of individual acts. Nozick does accord a couple of pages to the question of pollution, including that of when ‘*individuals have widespread and individually miniscule effects*’.<sup>20</sup> He regards the problem as redressable within the legal system, by means of group law suits. However, this assumes an institutional framework within which such suits can be made. It is precisely the lack of such a framework (say, according to which Bangladeshi farmers can sue US polluters) which is the problem in the global environmental case.

The Hegel-inspired “society of states” argument involves, as Caney notes, two claims: that the state is valuable because it is necessary to combat atomism and provide recognition and political unity (otherwise, people feel alienated), and that this instrumental value of the state is upheld only if that state is respected by other states, i.e. if it is a member in a society of independent, sovereign states.<sup>21</sup>

Borrowing Caney’s argument, we can reject this line of reasoning. Whilst it is plausible that humans require membership in a political regime to live flourishing lives (on Nussbaum’s approach, control over one’s political environment is a component of a central functioning<sup>22</sup>), that does not imply that those political regimes must be sovereign states with unlimited authority over all issues (that is, with comprehensiveness and supremacy). Rather, Caney argues:

---

<sup>20</sup> Nozick 1974, 80 [his emphasis]

<sup>21</sup> Caney 2007, 170-1, citing Mervyn Frost, *Ethics in International Relations* (Cambridge University Press, 1996)

<sup>22</sup> Nussbaum 2000b, 80

Surely what is needed is membership of political units with considerable autonomy which are recognised by other units to have worth and autonomy. And such institutions are quite compatible with the cosmopolitan framework defended early (i.e. one in which supra-state political institutions regulate and monitor the conduct of political communities).<sup>23</sup>

The other two statist arguments require more detailed consideration, and will form the subject of the next two sections.

#### **IV. Same benefits argument: is state co-operation enough?**

The same benefits argument denies the need for any global decision-making institution, and accordingly for the weakening of state supremacy outlined in Section II. The optimal result, it claims, will be achieved by the actions and co-operation of states.

This could be read two ways: *either* as the claim that states, acting independently, will behave in ways which, in combination, solve the problem of climate change, *or* as the more sophisticated objection that, by co-operating voluntarily with one another, states could retain all the current aspects of state sovereignty and still tackle climate change. This latter, again, could be read in two ways: as claiming that the combination of various agreements between different states (bilateral, trilateral or multilateral) would be sufficient to tackle climate change, or as asserting that a voluntarily upheld agreement, between *all* states, to co-operate in the aim of tackling climate change, would be adequate. In this last case, the challenge is not so much to the idea that we need to decide what to do at a global level (and so that individuals, albeit via states, should be promoting such collective action), as to the idea that such collective action must be implemented by a decision-making institution with coercive authority.

As an objection, this final version is the most plausible. Accordingly, I will focus on undermining it. However, in doing so, I will also respond to the weaker versions.

A preliminary point against the same benefits argument is that agreements between states will not be enough unless they have force over international

---

<sup>23</sup> Caney 2007, 171

corporations, who are significant players in the industries most impacting climate change.<sup>24</sup> That aside, however, there are two main reasons why such agreements might be insufficient. These are analogous to the arguments offered, with regard to *individual* motivation, in the last chapter.

Recall George: the middle-aged, rich, childless American. There is no prudential motivation for him to contribute to an effort to tackle climate change, although the collective responsibility for environmental harm of the global elite, including him, gives him a moral duty to do so. Moreover, as that duty is grounded negatively (in weak collective responsibility for harm) I suggested that it might be legitimately coercively imposed.

Is there a state counterpart to George? Given the increasing proportion of members of all states who are young enough to have an all things considered (ATC) interest in tackling climate change, as well as the emotional stake most members have in future generations of citizens, I do not think there is any state that can plausibly make the case that it does not have a prudential incentive to co-operate with efforts to tackle climate change.<sup>25</sup>

Note that this is not the same as claiming that the state *itself* has an ATC interest in tackling climate change: as discussed in Chapter Four, the state is assumed not to have interests over and above those of individual members. Rather, given the way the state is structured, and the interest most individual members have in environmental preservation, it should (although, as I will discuss below, does not necessarily) follow that the state would espouse the *goal* of limiting climate change, if only in the strictly limited sense of acknowledging that it would prefer there to be less environmental damage. (This preference can be understood in much the same way that each individual in the Prisoner's Dilemma would prefer to spend less time in prison.)<sup>26</sup>

---

<sup>24</sup> Caney draws attention to the various entities which could potentially be held responsible for pollution: individuals, corporations, states and international regimes or institutions. (2005, 752-6)

<sup>25</sup> This point could be made more straightforwardly by extending the account of Chapter Two to include future generations as members of the collectivity. However, that is beyond the scope of this argument.

<sup>26</sup> This is a complicated point, as the state includes some who are not in the actual collectivity with respect to environmental protection. Thus, the claim is not that this is a common goal or interest of the state in the sense that it is one of the interests through which individuals are members of the state. Rather, the point is that, if, *in practice*, states are required to take a view on this, it would make sense for them to espouse the goal of cutting climate change.



Why, then, cannot we rely on co-operation by states to tackle climate change? We need to return to the arguments of Chapter Four and their application, in Chapter Five, to the environmental case. All the arguments (Inefficiency Argument, Ignorance Argument, Game Theoretic Arguments, Partial Conflict Arguments, and Rational Altruist Arguments) were seen to be relevant when it comes to the position of *individuals* with respect to tackling climate change.<sup>27</sup> I suggest that they also apply when we assume the parties to be states rather than individuals. Given the objection under consideration, I will focus in particular on those which bring out the need for not only collective but also enforceable decision making.

The Ignorance Argument appeals to the possibility of individual denial, either of an ATC interest or of mutual dependence with certain others for satisfaction of that interest. In practice, there are two common versions of the ignorance case with respect to the environment. Firstly, those who deny the fact of climate change, deny that it is the result of human behaviour, or produce an inadequately supported assertion that some technological solution will be forthcoming (at the most extreme, that we will simply find another planet to move to). This, as I have already suggested, may be the result of adjustment of factual beliefs to avoid a clash in the set of beliefs and desires.<sup>28</sup> Secondly, those who do know the facts but, given that the clash between beliefs and desires is not brought home to them on a regular basis, succeed in ignoring them rather than facing up to the need for lifestyle adjustments.

Can this argument be evaded by considering states rather than individuals? It might seem that state governments, with access to the most up-to-date scientific information and knowledge that they are responsible for (and answerable to) the younger as well as the older generations of citizens, could not be this ignorant. However, states are made up of individual citizens, some of whom may either deny the facts or ignore them, voting accordingly. Many governments are also influenced by dominant interest groups, including international corporations with heavy carbon footprints. This renders it all too plausible that whatever view they ought, prudentially, to espouse, states could in practice act like Chapter Four's Head-in-the-Sand farmers.

---

<sup>27</sup> I made this point in Chapter Five, Section II.

<sup>28</sup> Papineau 1978, 168. See also Chapter Four, Section III.

Moreover, even if most states do acknowledge the situation, there remains a Game Theoretic motivation problem which explains not only the need for a collective decision-making body, but also powers of enforceability. Recall the Prisoner's Dilemma. The combined result of each individual doing what she is rationally motivated to do, *qua* individual agent, is that things are worse off for all individuals than if they had co-operated. However, a unenforceable agreement to co-operate doesn't work because, unless the payoffs have actually been changed, each individual has an incentive to cheat.

States are in this situation with respect to cutting carbon emissions to tackle climate change. As regards the larger states at least, this is a multi-player Prisoner's Dilemma (albeit one in which, unlike with the Common Land Case, there isn't the additional complication of an incentive to free ride<sup>29</sup>).

#### *Multi-player Prisoner's Dilemma*

|         |                     |               |                     |
|---------|---------------------|---------------|---------------------|
|         |                     | Other states  | Other states        |
|         |                     | Cut emissions | Don't cut emissions |
| State A | Cut emissions       | S,S           | W,B                 |
| State A | Don't cut emissions | B,W           | T,T                 |

Thus, it seems, an unenforceable agreement between states, to co-operate, would not be enough.

However, there is a counterargument to this line of reasoning, brought out by the analogy with the Hobbesian state of nature.<sup>30</sup> Modern Hobbes theorists have made sense of an apparent clash between two "reason-based" lines of action in the state of nature (fight or seek peace), by modelling the situation as a Prisoner's Dilemma, with its clash between individual and collective rationality. However,

<sup>29</sup> Chapter Four, Section IV. Each *individual* has a rational incentive to free ride on the environmental efforts of others, but this is not the case with all states. As with individuals (but more obviously), there are some things all states must *not* do if the planet is to be preserved. Recall the distinction brought out in the last chapter between co-operating by omission and in a more positive sense, in this case by cutting down carbon emissions. With regard to the latter, it is possible that a smaller states could free ride on the environmental policies of big states and other small ones. However, in combination, even smaller states failing to co-operate would undermine the project. Moreover, bigger states are necessary for the success of the project, and so the free riding incentive does not apply to them.

<sup>30</sup> Chapter Five, Section IV. Note that Hobbes himself describes the international situation as a state of nature, although he, living in a generation where the impact of humans on the environment was relatively limited, does not see this as threatening our fundamental interests. (1651, 90)

as Jean Hampton points out, the incentive structure is changed in a repeated game.<sup>31</sup>

[I]n this case the parties to the bargain in the natural state know that they will have occasion to contract with each other in the future, and although they know that the action of breach is rational in the short run (i.e., in the first game), they know that in the long run the breach will deprive them of benefits from future bargains by creating distrust between them and putting the action of breach in equilibrium. Hence, the argument is that because contractual activity between people in the state of nature is likely to be frequent and open-ended, the keeping of a contract is always rational, because in addition to supplying each with the benefits of a particular bargain, it acts as a signal to each party that the other party will keep contracts in the future and thus enables both to reap the long-term profits of constant contractual activity.<sup>32</sup>

In other words, the payoffs change in a repeated game. Given that the environmental “game” is not a one-off Prisoner’s Dilemma, but an ongoing issue where there are likely to be repeat attempts at agreements, the corresponding argument would be that it would be rational for each individual state to sign up to, and stick to, emissions quotas.

That this has *not* been the case in practice hardly needs pointing out.<sup>33</sup> Moreover, drawing further on Hampton, it is possible to understand why.<sup>34</sup> She argues that conflict arises in the Hobbesian state of nature, despite the rationality of making and sticking to contracts, because of the parties’ short-sighted pursuit of self-preservation.

Many people fail to appreciate the long-term benefits of cooperation and opt instead for the short-term benefits of noncooperation, and the rest are legitimately fearful enough of this short-sightedness afflicting their partners to doubt that cooperation would have any educative effects. This worry could then force even a farsighted person to take a single-play orientation, with the result that the uncooperative action would dominate.<sup>35</sup>

---

<sup>31</sup> Hampton 1986, 74-9

<sup>32</sup> *Ibid.*, 75

<sup>33</sup> See Gardiner 2004, 589-95 for a discussion of the failure of the Kyoto Protocol.

<sup>34</sup> Hampton 1986, 80-88

<sup>35</sup> *Ibid.*, 81

There are, Hampton suggests, two reasons for short-sightedness. Firstly, some ‘less intellectually talented inhabitants’<sup>36</sup> will fail to understand the iterated Prisoner’s Dilemma situation. Either they do not grasp that they are in a multi-play situation, or they fail to appreciate the advantages of long-term co-operation. This is precisely the point raised above, in the Ignorance Argument. Certain states, with some voters in the situation of George, some others thinking like Young George, and big business holding huge political influence, might be constrained not to acknowledge the need for long-term co-operation. Modern democratic politics, by its nature, is a short-term business: governments have an interest in prioritising policies which have their payback before the next election, and future generations don’t have a vote.

Hampton’s second explanation of short-sightedness in the Hobbesian state of nature is that ‘the prospect of future contractual interactions with... present partners is too remote to warrant [co-operating in the present game]’. This, however, is rendered plausible in the state of nature only because individuals are supposed to be ‘independent of *all* society’.<sup>37</sup> It is less so with individuals in the actual world, and considerably less so again when it comes to states, who are limited in number, and have existing channels of communication.

However, the first point remains convincing. Moreover, it is not necessary for the argument to work that all states would actually be short-sighted, only that it is sufficiently likely enough that some would be short-sighted (as they have proved to be in practice) for fear of being cheated on to motivate the others not to comply with any non-enforceable agreement.

Thus, any interstate agreement to tackle climate change (say, by setting emissions levels for each state) would require enforceability by the other states – effectively, by the overall collectivity. This would, in all but name, be the kind of global institution defended here.<sup>38</sup>

---

<sup>36</sup> Ibid., 82

<sup>37</sup> Ibid., 82

<sup>38</sup> This raises the related question of how, if we are in a situation analogous to the Hobbesian state of nature, agreement could ever be reached (and stuck to) to establish such a global institution. Again, it is helpful to borrow from Hampton. She suggests that the process of accepting a sovereign can be a Co-ordination Problem rather than a Prisoner’s Dilemma, because once there is a potential sovereign on the table, each player would be worse off refusing to co-operate if others did, as the balance of power would be against them. (Hampton 1986, 138-50) Similarly, if the majority of states (those that have shown themselves to be farsighted) were to establish an environmental collective decision-making body, give other states the chance to comply (and be

However, even if the same result could be achieved by various bi- and multilateral agreements, there is another reason for requiring a formal, global institution with the power to oversee the allocation of targets. This borrows from argument given by Rawls, in the domestic case, against the libertarian, and draws attention to a weakness in Rawls' own international account.<sup>39</sup>

On the libertarian account, the only just outcome is that resulting from just (free) transfers from an initially just acquisition. There is no need for more than a minimal state (one limited to 'the narrow functions of protection against force, theft, fraud, enforcement of contracts, and so on'<sup>40</sup>). Rawls, in defending his principles of justice for the basic structure of society, appeals to the possibility of an overall unjust outcome, resulting from a combination of individually just transactions.

[S]uppose we begin with the initially attractive idea that social circumstances and people's relationships to one another should develop over time in accordance with free agreements fairly arrived at and fully honoured. Straightaway we need an account of when agreements are free and the social circumstances under which they are reaching are fair. In addition, while these conditions may be fair at an earlier time, the accumulated results of many separate and ostensibly fair agreements, together with social trends and historical contingencies, are likely in the course of time to alter citizens' relationships and opportunities so that the conditions for free and fair agreements no longer hold.<sup>41</sup>

The idea is that the basic structure establishes a framework within which individuals can interact voluntarily, without thereby undermining the justice of the overall process.

Exactly the same point can be made at the global level, although Rawls does not acknowledge this.<sup>42</sup> Given that the world is, environmentally speaking, as

---

represented on the body), but threaten sanctions if they do not, this could be a way to change the payoffs so that all states complied.

<sup>39</sup> Rawls 1993, 262-9

<sup>40</sup> Ibid., 262

<sup>41</sup> Ibid., 265-6

<sup>42</sup> At the international level, Rawls leaves everything to transactions between supposedly free and equal liberal (or at least decent) peoples. His failure to consider the possibility of the kind of structure advocated in Section II can be attributed, I suggest, to two ways in which he appears to underestimate the complexity of the debate. Firstly, he offers a vehement rebuttal of the idea of a world state (1999, 36) but, like the proponent of the liberty-based argument, above, appears to neglect the possibility of some institutional set-up which is neither a world state nor a "society of peoples". (1999, 3) Secondly, in dismissing the suggestion that natural resources should be redistributed between countries on the grounds that the crucial element in how a country fares is

closed as Rawls' paradigm society, the outcome of "free and equal" treaties and transactions between states may undermine the capacity of other states (for example, low lying, vulnerable countries such as Bangladesh) to endure as a liberal or decent state. Exactly as Rawls notes of the domestic case, so we can say of the international one:

[T]he conditions necessary for background justice can be undermined, even though nobody acts unfairly or is aware of how the overall result of many separate exchanges affects the opportunities of others. There are no feasible rules that it is practicable to require economic agents to follow in their day-to-day transactions that can prevent these undesirable consequences. These consequences are often so far in the future, or so indirect, that the attempt to forestall them by restrictive rules that apply to individuals would be an excessive if not an impossible burden.<sup>43</sup>

The argument here is limited to the need for global *environmental* decision-making institutions, although it is plausible that the Rawlsian argument could be used against him with respect to distributive justice as well.

## V. Moral burdens argument

This section will consider the second statist objection to the model in Sections I and II: the *moral burdens* argument. This claims that the model, in obliging individuals to consider the interests of those geographically or temporally distant, places too heavy a burden on them. As I will argue, as a statist objection, it can be rejected. However, it raises questions which impact not just on the claims of this chapter, but on the theoretical framework on which they were based. I will also address such points, although a full discussion will have to be reserved for later projects.

The appeal to burdensomeness can be read in three ways: as requiring too much of the individual in terms of knowledge of the far-reaching impact of her actions, as requiring an individual to make unreasonable sacrifices of her own

---

its political culture, not the level of resources, Rawls appears completely to ignore the numerous negative externalities attached to the *use* of those natural resources.

<sup>43</sup> Rawls 1999, 266

goals and interests, or as requiring an individual to prioritise distantly-grounded duties over the local moral claims of, for example, family or community.

The first of these is quickly addressed: establishing an institutional structure, within which common interests are met and collective harms mitigated, should actually make life *easier* for individuals. The framework established by collective decision making at each appropriate level would leave individuals free to carry out their own lives within the sphere of the restrictions handed down to them, knowing that so long as they complied with those restrictions they would not have to worry about their collectively acquired responsibilities. This is analogous to the Rawlsian basic structure, as discussed in the last section, which is intended to leave individuals free to exercise their freedom and partiality, within a set of constraints which ensure that justice is satisfied.

The rest of this section will focus on the second two readings of the objection. These could be put against both main lines of argument in this thesis. That is, to the defence of collective and correlative individual duties by appeal via the non-intentionalist account of collectivities, to prudential considerations backed by considerations of fairness, their defence on the basis of negative moral responsibility for harm.

Let us begin with the former. The problem is that an individual's environmental responsibilities, acquired through membership of the global collectivity, could clash with pursuit of her own individual goals or interests, *or* with duties she acquires through membership in other collectivities. For example, Young George's own goals involve living a high carbon lifestyle, which clashes with environmental duties. Alternatively, someone of my generation might also be an executive in an oil company, rendering her part of the global collectivity with respect to environmental protection, but also part of a smaller collectivity with rather different shared goals.

In general, the possibility of clashing duties as a result of membership of different collectivities, or a clash between individual goals and duties acquired through membership of a collectivity, need not pose insuperable problems. As Chapter Four, Section X pointed out, consideration of other interests or goals may mean that the individual, *all things considered*, would rather not be a member of some collectivity. She is not then generally bound to act as though she were. Recall Jane, the potential car sharer. Alternatively, consider an

individual who is a member of both the local mountain biking club and a local wildlife preservation society. Suppose the preservation of a rare bird required the closing of a particular trail to mountain bikes. The goals would clash. However, it would be up to the individual which she prioritised: she would not be bound to remain in the mountain biking society, with its goal of fighting to keep the trail open, if she regarded the preservation of the birds as more important.

Such opt-out clauses have been ruled out only in the case of certain collectivities: those where individuals are mutually dependent for the satisfaction of their ATC interests. This chapter has assigned priority to decisions made in those collectivities and, within the set of such collectivities, to those decisions made at the wider collective level: that is, those where the co-operation of most people is required. The two most obvious cases are global environmental decisions, and those relating to political questions at state level.

This can be defended against the burdensomeness complaint as follows. Firstly, if individuals *are*, as a matter of fact, in a situation such that satisfaction of their fundamental interests depends on co-operation at the global level, it is not very helpful to argue that having to take account of this, and prioritise decisions made there, renders their lives more difficult. General failure to do so will make their lives *worse*. This is so even if fundamental interests are also at stake in the smaller collectivity, as with the state, or if our oil company executive's job depended on the industry's continuing to expand. If the problem of climate change is not addressed – and I have argued that it can only successfully be tackled at global level – then, no matter what is done at the level of smaller collectivities, each individual's fundamental interests will be at risk. It therefore makes sense for state (and ultimately individual) decisions to be made within the boundaries drawn at the global level.

Lest this seem paternalistic in terms of practical application (given that not all individuals *do* accept the prudential reasoning), it can be supplemented with a second point. There is a distinction between this environmental case and that in which an individual risks *only his own* central capabilities in pursuit of some other goal. While it could be said to be paternalistic to interfere, purely on the ground of his own long-term interests, with someone who risks a central



capability to pursue some ideological goal or to keep a job,<sup>44</sup> in this instance the fundamental interests of all the others in the collectivity (near *and* far) are also at stake. This, as has been argued at length in the previous chapter, makes a *morally* significant difference.<sup>45</sup>

Consider now the morally salient potential collectivity constituted by those individuals collectively responsible for environmental harm. Here, I have contended, there is a negatively grounded collective duty (and corresponding individual duties) to establish and abide by the global decision-making body. Given this, the moral burdens objection is already weakened: it is harder to slip out of a responsibility not to *harm* someone else, just because that person is further away, than to evade the more controversial responsibility to *help* them.

Moreover, this way of tackling the issue should make things less rather than more burdensome for individuals. Chapter One made the point that the responsibility in this case is collective. Once the institutional structure envisaged in Section II is in place, no individual is required to negate the collective harm alone: each has only to play her part in a collective attempt to do so.

Nothing has been said about how collective institutions would allocate duties to individuals. However, recalling the moral “excuses” of the last chapter, it should be clear that excessive burdensomeness could exclude an individual from having to comply. This is something the environmental body would have to take into account in allocating duties. Thus, for example, the owner of a small but polluting factory in a developing country, whose own livelihood or that of his family depended on the business, would not be required to cut emissions in the same way as a major US player would.<sup>46</sup>

A full discussion of this topic is beyond the scope of this project.<sup>47</sup> My point, for now, is that the imposition of certain restrictions at global level, to remedy

---

<sup>44</sup> Although we can certainly criticise the state framework in which he is obliged to make such a choice, and, as Chapter Three made clear, we can say he is undermining his own flourishing by so acting.

<sup>45</sup> Chapter Five, Sections III, IV and VIII

<sup>46</sup> As Shue puts it, at the interstate level: ‘[W]hatever justice may positively require, it does not permit that poor nations be told to sell *their* blankets in order that rich nations may keep *their* jewellery.’ (1992, 397)

<sup>47</sup> The question of further conditions on the legitimacy of a global environmental decision-making body will be part of my post doctoral research. These conditions would include the fairness of allocation of duties.

collective environmental harm, need not mean depriving individuals of scope to pursue individual projects, and to satisfy their moral duties closer to home.

This should be enough to respond to the moral burdens argument, as a statist objection to the institutional framework laid out in Section II. However, a related question does arise. I have defended individual duties, derived from collective responsibility for harm, to *promote* institutions (where they do not already exist) to address the harm. Moreover, I suggested supplementary individual duties, to mimic such duties as would be assigned by a collective decision-making process, or to attempt to mitigate the harm directly.<sup>48</sup>

These also give rise to concerns about burdensomeness. Promotional and supplementary duties could clash with individuals' pursuit of their own goals, with individuals' moral obligations closer to home (for example, their duties to their families), or even with each other.

Further discussion is needed of how far individual duties to promote collective institutions should go (when, for example, an individual might instead legitimately opt out of the harming group) and under exactly which circumstances supplementary duties come into play. This would naturally take account of moral excuses parallel to those discussed in Chapter Five. These include cost to the individual, or cost in terms of preventing worse alternatives, of fulfilling the duties.

Some preliminary remarks were made in Chapter One and I plan to return to it in later work, but a thorough discussion of this issue is beyond the scope of the current project.<sup>49</sup> The potential for conflicting moral duties, including the potential for tragic moral clashes, where an individual is stuck between two morally bad options, is not unique to my model. I will, accordingly, limit myself to making a few points in defence of this thesis, and in particular to eliminating an especially worrying potential charge: that I have been self-contradictory.

This latter attack arises as follows. Recall that, in Chapter One, I motivated my expansion of collective responsibility in part by appeal to Scheffler, who highlighted the difficulty of addressing some of the most pressing current global problems with only the concept of *individual* responsibility. Rather than abandon the intuitive pull of a common sense morality at the individual level, which

---

<sup>48</sup> Chapter One, Section IV

<sup>49</sup> This forms part of my post doctoral research proposal

prioritised negative over positive duties and duties to those closer to us over those further away, it was suggested that the most constructive approach would be to focus on collective responsibility. Now, given that I have argued that an *individual*, as a result of being part of a group collectively responsible in a weak sense for harm, can acquire certain duties, it might seem that I am back where I started: with counterintuitive results.

This is not so. Firstly, it is important to be clear about the line of argument. As specified in Chapter One, the point was never to deny individual responsibilities with regard to such far-reaching or combined effects altogether, but rather to channel them in an appropriate way. The idea was to expand the options for an individual, faced with the reality of the situation, beyond the following, one of which has practically disastrous consequences, and the other of which is too restrictive: do a certain action on the basis that a) you can't be expected to see how that exact act will affect others within the far-reaching global network, and b) your act alone won't make any difference anyway; *or* refrain from the action altogether because if everyone acts in the same way, it will bring about harm.

Rather, I have argued that both collectively caused harm, and the possibility of collective action to remedy that harm, should be taken into account in determining how the group (collectively), and each individual within it, should behave. Thus, while individuals may well be expected to behave in certain ways because of collective responsibility for harm, the duties they acquire are derived on a different level to those acquired through their individual acts, taken in isolation. (That is not to say, however, that one or the other must always take priority.)

With this in mind, the following defence can be made against the charge that individual promotional and supplementary duties, acquired through collective responsibility for harm, impose too heavy a burden. Firstly, it is less burdensome if an individual is expected to take into account the fact that the predictable combination of acts, including hers, will have a negative impact, than if she is expected to understand the exact impact of *her individual act* within the 'various large-scale causal processes and patterns of activity'<sup>50</sup> within which it is subsumed.

---

<sup>50</sup> Scheffler 2001, 43

Secondly, with regard to the potential for clash between promotional and supplementary duties, Chapter One stated that the former are the primary duties incurred by individuals in cases of weak collective responsibility for harm. Supplementary duties come into force if there is little chance of the promotional efforts succeeding or, if compatible, alongside them.<sup>51</sup>

Thirdly, relatedly, my discussion of weak collective responsibility indicated that individual mimicking duties would *not* be morally required if fulfilling them would make no difference.<sup>52</sup> Thus, the model can (as is practically essential) accommodate the far reaching combined impact of our actions, without requiring us, as individuals, to make pointless moral sacrifices.

## **VI. Past harmers objection**

Having tackled the statist objections and acknowledged their wider implications, I turn now to an argument which focuses on the collective harm aspect of my defence of environmental duties.

Caney, debating the fairest way to allocate the burdens of climate change, rejects what he calls the Polluter Pays Principal, in part by appeal to the significant proportion of the harm done by previous generations.<sup>53</sup> Might a parallel objection be raised to my reliance on collective responsibility of current generations for environmental harm in deriving collective (and correlative individual) duties to tackle climate change?

My response is as follows. Firstly, we *are* currently harming, collectively. Thus, at the very least, we have a collective responsibility to end that harm. This in itself would be enough to make the case for the global decision-making body, and for individual duties to promote such an institution.

A second possible reply would go as follows. Recall the discussion of Chapter One, Section IV. There, it was suggested that, however plausible Murphy's claim that 'a principle of beneficence should not increase its demands on agents as expected compliance with the principle by other agents decreases',<sup>54</sup> a parallel

---

<sup>51</sup> Chapter One, Section IV

<sup>52</sup> Chapter One, Section IV

<sup>53</sup> Caney 2005, 752-62

<sup>54</sup> Murphy 1993, 278

claim regarding the individual duties incurred through collective responsibility for harm would be less obviously acceptable. If the group of those collectively responsible for environmental harm also includes those who, by virtue of being dead, have defaulted on their duties, then, although a morally non-ideal solution, it could be that the others in the group should bear the burdens (especially given that they began to contribute against a known background of harm already done). However, this line of argument would require considerable further work.

Thirdly, there remains the other side of my account. This objection could be used to bring out the importance of combining both lines of argument: prudentially grounded duties, based on collectivityhood, and negatively morally grounded duties, based on membership of a morally salient potential collectivity. It is in most of our ATC interest not only to stop adding to climate change, but to find a way of tackling the harm already done. This supplements the case for a global decision-making body, and extends its mandate.

Finally, relatedly, it is the collective body that would *then* have the responsibility for allocating duties to individuals within the collectivity or morally salient potential collectivity. The task of this thesis has been importantly different from that of Caney's book. The question of how such duties could fairly be allocated, which would presumably include considerations of contribution levels, as well of who benefits, or (as Caney advocates) who can bear the costs, is the natural next stage for discussion.

## **VII. Too inclusive objection**

Finally, consider the claim that the account defended in this thesis has not only the implications I have defended at the global level, but other, less intuitive implications: it is too inclusive. This objection might be levelled at both elements of the case for global environmental decision making, and I will respond by considering each in turn.

Begin with the claim that most of us constitute a collectivity through our ATC interest in environmental preservation. The worry is that if we can derive a global (or near global) collectivity and a corresponding prudential incentive to adopt a global decision-making body by appeal to some (not necessarily acknowledged) interest, we could be opening the door to all sorts of global authorities. Consider

a couple of counterintuitive examples. We all have an interest in staying fit and healthy. Does this mean we should adopt a global fitness institution, mandated to force us to exercise? The preservation of great art is desired by many, in countries all across the world. Should we therefore have a global Department for the Arts, with authority to dictate to national and sub national galleries and museums?<sup>55</sup>

In both cases, the answer is no. Nor does my model imply otherwise. A version of this objection was already rejected in Chapter Two, Section X. Individuals can only become members of a collectivity through: a genuinely (acknowledged) shared goal, selfish goals which require co-operation for their achievement (and this is outweighed, as discussed in Chapter Four, Section X, if individuals would rather give up the goal than co-operate with the relevant others to achieve it), or an ATC interest. The only ATC interest defended in this thesis, as applying acknowledged or not, was that each one of us has in enjoying the central human capabilities. Moreover, individuals have to be mutually dependent for the satisfaction of their ATC interest: it has to be something which can only be satisfied by all, for all.

These conditions exclude the cases above. While the opportunity for a healthy life is a central *capability*, this does not legitimate forcing individuals to exercise the *functioning* (health or, in this case, fitness, which is only one aspect of the functioning). Even if keeping fit could be said to be in the interest of each and every one of us, we are not mutually dependent globally for the achievement of this end. It can be pursued individually, or at least through various local, voluntary collectivities.

The preservation of great art is even less obviously in each individual's interest, and even if it were, it does not require pursuit at a global level. The fact that many individuals do have an interest in great art, and that they do not share the same state, is not enough in itself to require supra-state co-ordination. Unlike environmental preservation, it seems to be something that can be left to the art industry within each state to sort out. (Of course, a case might be made for international co-operation, or even some international court to adjudicate with respect to art *theft*, but that is not quite the same point.)

---

<sup>55</sup> Thanks to Jonathan Wolff for the examples.

The arguments of the last chapters supplement the response to this objection. The case for prudentially-grounded collective decision making has been restricted to scenarios where there is a common interest or goal in the sense necessary for collectivityhood, *and* there is a need, according to the kind of argument offered in Chapter Four, for *collective* action. Chapters Four and Five have, moreover, made it clear that further arguments are necessary, even in such cases, before the possibility of legitimate coercive enforcement of collective decisions can be admitted.

Concerns might also be raised about inconclusiveness with regard to the other side of the defence of global environmental duties: collective responsibility for harm. Weak collective responsibility is a very broad notion. I have already had to defend it against objections in Chapter One.<sup>56</sup> Now, the question is: if we can be collectively responsible, in a sense giving rise to all these collective and correlative individual duties, for environmental damage, how else might we acquire such duties? Consider the following cases.

Firstly, suppose a number of walkers, independently, venture onto an area of apparently solid ground and their combined weight causes a landslide which seriously injures other walkers below. Are they collectively morally responsible? Secondly, suppose the combined sound of schoolchildren, chattering in their twos and threes as they pass my window for ten minutes or so every morning, drives me into a frenzy of irritation. Does this give them a duty to co-ordinate a rota so that only one or two speaks at a time, or some take another route? Finally, if a group of beautiful girls move to a small village and all the young men rush to marry them, are they (the beautiful girls and the susceptible men) collectively responsible for morally significant harm to the girls already in the village, whom the men would otherwise have married?<sup>57</sup> Do they acquire collective and correlative individual corrective duties as a result? Are the beautiful girls morally required to give up their partners, or at least ship in some other men for the other women?

Again, the answer is no. These examples overlook the various restrictions imposed by the model, both in terms of the original conditions for weak

---

<sup>56</sup> Chapter One, Sections IV and V

<sup>57</sup> A variant of Nozick's example of Z and Z' who are left with no choice but to marry each other, repulsive though they find each other, as the more attractive A to Y and A' to Y' have already paired off. (1974, 263)

collective responsibility, and in terms of the moral “excuses” which might be allowed to render harm morally permissible.

Recall the two-part sufficient condition for weak collective responsibility, introduced in Chapter One, Section III:

- the individuals acted in ways which, in aggregate, caused harm, and which they were aware (or could reasonably be expected to have foreseen) would, in aggregate, cause harm (although each only intentionally performed his own act); and
- they were all aware (or could reasonably be expected to have foreseen) that there were enough others similarly placed (and so similarly motivated to act) for the combined actions to bring about the harm.

This rules out such one-off, unpredictable cases as that of the walkers.

Moreover, it has been understood throughout that not all negative impact on another person constitutes morally significant harm. While it has not been claimed positively that nothing else counts as harm in the relevant sense, the only impact on another which it *has* been asserted should count as morally relevant harm is the undermining of a fundamental interest or capability. This rules out the schoolchildren case, although not the marriage one (assuming the village is relatively isolated, and this could be the women’s only chance to form a loving relationship and have children).

However, the moral “excuses” outlined in Chapter Five included the cost to oneself of not doing the harming action.<sup>58</sup> This applies in the marriage case, where the cost to the men and the beautiful women would be being deprived of choice in their life partner. (Appeal to some uncontroversial individual *rights*, notably rights of self ownership, would also be enough to refute any attempt to coercively enforce any demand, even if such a duty could have been convincingly defended, for the men to marry the less attractive women.)

Some controversial cases might, however, be included by my model. For example, Kernohan highlights cases of cultural oppression, where (in various ways, such as through pornography) a negative impression is perpetuated of a

---

<sup>58</sup> Chapter Five, Section VI



certain group within society, which leads its members to see themselves as of unequal moral worth.<sup>59</sup> He defends his Accumulative Harm Principle with reference to such examples. Discussion of this kind of cultural oppression is beyond the scope of this thesis but, should it turn out to be a case of weak collective responsibility for harm, on the model defended in Chapter Two, and not be ruled out by the excuses of Chapter Five, I would simply bite the bullet and accept the implications in terms of collective (and correlative individual) duties.

I have also left open the possibility that certain global duties of distributive justice could be justified by appeal to the collective harm done, economically, by rich westerners (and corporations and governments) to the global poor. Rather than ground such duties, as Pogge does, in our upholding an institutional framework within which certain individuals' human rights are not met, the normative point would be that it is the predictable result of the combination of our individual choices (as consumers, among other things) that many in developing countries are restricted to a lifestyle without secure access to central capabilities.<sup>60</sup> This, as I said in Section III, goes beyond the scope of my project. However, as the cosmopolitan criticism highlighted there makes clear, it is far from obvious that such an implication, were it to be demonstrated, would be a *disadvantage* for my account, or show it to be implausibly inclusive.

## Conclusion

This chapter has drawn together the arguments of the rest of the thesis, to argue that decisions related to environmental protection *should* (prudentially and morally) be made at a global level. These decisions would then be handed down as a framework within which nation states (and other supra-state institutions or federations) can exercise their political authority. It is further suggested that,

---

<sup>59</sup> Kernohan 1998, 6-19

<sup>60</sup> Pogge 2002, 45-8. The two questions are already linked by Shue, who argues that a global climate change policy should not only require richer nations to bear more of the costs of prevention, but also to assist poorer nations in coping with such climate change as is now unpreventable. He argues: 'The rich nations have indisputably so far caused most of whatever problems there are in the cases of ozone depletion and global warming. Consequently, any attempt by rich nations to wash their hands of any resultant harms that are not prevented would be doubly unfair.' (1992, 391)

those nation states having made their legislative, executive and judicial decisions, sub-state communities should be left with space in which to pursue their common goals.

To make such environmental decisions globally will require some kind of global institution. This is exactly the kind of suggestion designed to raise western hackles. However, four statist arguments have been rejected, rebutting a challenge to the Decision-Making Claim and the assignment of authority to political and social structures sketched on the back of it. Objections to the overall model have also been acknowledged, and largely rejected.

## Conclusion

This thesis has argued for global institutional change: in particular, it has advocated a global environmental decision-making body.

I have defended a collective moral duty to tackle climate change, derived from our collective moral responsibility, in a weak but significant sense, for environmental harm, both to future generations and to those (primarily the global poor) whose central capabilities are already threatened by environmental damage. I have also defended a collective prudential incentive to do so, by appeal to the idea that most humans are members of a global collectivity. The achievement of this morally required, collectively rational goal requires us, I have argued, to make the relevant decisions *qua* collectivity (i.e. at the global level) and hand them down as restrictions on individual (and state) action. In other words, a global decision-making structure is required.

I have argued that powers of enforcement over non-complying individuals (or states) will also be necessary. Chapter Five made a significant step towards *justifying* such coercive authority, by appeal to a collectivised version of Mill's Harm Principle. It also offered a more tentative case for legitimate collective coercion, based on a modified version of Hart's Principle of Fairness, supplemented by appeal to positive moral claims.

The proposed changes to the global institutional framework have been defended against four statist objections, and the arguments of the overall thesis defended against two further attacks. The fact, brought out by Caney, that much environmental harm was done by earlier generations, serves, it was argued, not to undermine my collective harm-based argument, but only to reinforce the importance of both sides of my model. Moreover, I have argued that the account defended here is not rendered weak by its inclusivity. It may have some controversial implications, but not clearly counterintuitive ones.

However, there are limits to what it has been possible to achieve. Only one side of the Decision-Making Claim of Chapter Six was defended in any detail. The other side gives rise to a familiar communitarian argument: that too much authority has been ceded from the state to substate collectivities. Conversely, cosmopolitans about global justice would be likely to complain that the transfer

of power *upwards* does not go far enough: that the global institution should enforce global duties of distributive justice. Both these points are beyond the scope of this thesis, although they are interesting topics in themselves and it is plausible that my model could be extended into the global justice debate.

More importantly, in terms of the aims of this specific project, two key points call for further debate. Firstly, Chapter One defended an individual duty to promote the necessary collective structures to address weak collective harm, and suggested two supplementary individual duties: to mimic the duties such an institution would assign, or to attempt to mitigate the harm. As Chapter Six acknowledged, in the environmental case, these could impose considerable costs on individuals. They could be burdensome in terms of clashes with both individual projects and moral duties acquired as individuals or members of smaller collectivities. Thus, questions were raised as to when, and to what extent, an individual would be morally obliged to fulfil her collectively-derived duties. These questions were left open.

Secondly, more work is needed on whether a conclusive defence of the coercive enforcement of environmental duties (in general) is available through the line of argument sketched in Chapter Five, Sections VII and VIII. Relatedly, questions arise as to what criteria any particular global environmental institution would have to meet before it could be said to have legitimate authority. Such criteria are likely to include conditions regarding both the representativeness of the decision-making process (or at least, perhaps, its ensuring that all have the *opportunity* to be represented), and the fairness of the distribution of burdens.

Two further issues, barely touched on in this project, also naturally suggest themselves. I have talked very little about the issue of population growth, but this is a key contributor to climate change and, as such, would presumably need to be addressed by any global environmental decision-making structure. However, any regulation is likely to provoke emotive objections, and appeal to an individual “right to procreate”.<sup>1</sup> Moreover, I have done no more than hint that the model could be expanded to include (at least) sentient non-human animals, whose flourishing can be said to have moral value and who could, using an expansion of the capabilities approach, be said to have certain fundamental interests.

---

<sup>1</sup> See also Chapter One, Footnote 50 for a specific problem relating to overpopulation and weak collective responsibility.

I hope, in my next research project, to give all four of these points the attention they merit. This thesis could only ever be a start. However, it has covered considerable ground. I have outlined a normative framework within which to tackle some of our most pressing global problems, and indicated how it might be applied to what is arguably the most pressing of them all. I hope that this, in itself, is both philosophically interesting and practically relevant.

## Bibliography

- Alkire, Sabina (2002): *Valuing Freedoms: Sen's Capability Approach and Poverty Reduction* (Oxford University Press)
- Agarwal, Anil and Narain, Sunita (1991): *Global Warming in an Unequal World* (Centre for Science and Environment)
- Austin, J L (1956): "A Plea for Excuses" (*Proceedings of the Aristotelian Society*, New Series 57)
- Brandt, Richard B (1979): *A Theory of the Good and the Right* (Oxford University Press)
- Caney, Simon (2005): "Cosmopolitan Justice, Responsibility and Global Climate Change" (*Leiden Journal of International Law* 18)
- Caney, Simon (2007): *Justice Beyond Borders* (Oxford University Press)
- Casal, Paula (1999): "Environmentalism, Procreation, and the Principle of Fairness" (*Public Affairs Quarterly* 13)
- Cohen, G A (1988): *History, Labour and Freedom: Themes from Marx* (Oxford University Press)
- Cohen, G A (1995): *Self-Ownership, Freedom and Equality* (Cambridge University Press)
- Cohen, Joshua (1995): Review of *Inequality Reexamined* (*Journal of Philosophy* 92)
- Cripps, Elizabeth Blanche (2005): *Individuals and Society: A Non-Intentionalist Account of Collectivities and Social Groups* (UCL MPhil thesis)
- Cripps, Elizabeth Blanche (unpublished): "The Discursive Dilemma and Collective Decision-Making" (presented at the Institute of Philosophy Graduate Conference, Senate House, London, 10 February 2006)
- Dworkin, Ronald (2000): *Sovereign Virtue: The Theory and Practice of Equality* (Harvard University Press)
- Feinberg, Joel (1987): *The Moral Limits of the Criminal Law Volume I: Harm to Others* (Oxford University Press)
- Føllesdal, Andreas (1998): "Survey Article: Subsidiarity" (*Journal of Political Philosophy* 6)
- Foot, Philippa (1978a): "The Problem of Abortion and Negative and Positive Duty: A Reply to James LeRoy Smith" (*The Journal of Medicine and Philosophy* 3)

- Foot, Philippa (1978b): *Virtues and Vices* (Oxford University Press, 2002 edition)
- Foot, Philippa (2001): *Natural Goodness* (Oxford University Press)
- Gardiner, Stephen (2002): "The Real Tragedy of the Commons" (*Philosophy and Public Affairs* 30)
- Gardiner, Stephen (2004): "Survey Article: Ethics and Global Climate Change" (*Ethics* 114)
- Gilbert, Margaret (1989): *On Social Facts* (Routledge)
- Gilbert, Margaret (2006): *A Theory of Political Obligation* (Oxford University Press)
- Graham, Keith (2002): *Practical Reasoning in a Social World* (Cambridge University Press)
- Hardin, Garrett (1968): "The Tragedy of the Commons" (*Science* 162)
- Hamburger, Henry (1979): *Games as Models of Social Phenomena* (W H Freeman)
- Hampton, Jean (1986): *Hobbes and the Social Contract Tradition* (Cambridge University Press)
- Hart, H L A (1955): "Are There Any Natural Rights?" (*Philosophical Review* 64)
- Hobbes, Thomas (1651): *Leviathan* (Revised student edition, ed. Richard Tuck, Cambridge University Press, 1996)
- Hume, David (1777): *An Enquiry Concerning the Principles of Morals* (*Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, David Hume, ed. L A Selby-Bigge, third edition with revisions by P H Nidditch, Oxford University Press, 1975)
- Hurley, S L (1989): *Natural Reasons: Personality and Polity* (Oxford University Press)
- Husak, Douglas (2005): "On the Supposed Priority of Justification to Excuse" (*Law and Philosophy* 24)
- IPCC (Intergovernmental Panel on Climate Change) (2007): *Climate Change 2007: Synthesis Report* (Cambridge University Press; available from [www.ipcc.ch](http://www.ipcc.ch))
- Kant, Immanuel (1785): *Groundwork of the Metaphysics of Morals* (ed. Mary Gregor, Cambridge University Press, 1998)
- Kernohan, Andrew (1998): *Liberalism, Equality, and Cultural Oppression* (Cambridge University Press)
- Klosko, George (1987): "The Principle of Fairness and Political Obligation" (*Ethics* 97)

LeRoy Smith, James (1978): "The Problem of Abortion and Negative and Positive Duty" (*The Journal of Medicine and Philosophy* 3)

Lichtenberg, Judith (1981): "National Boundaries and Moral Boundaries: A Cosmopolitan View" (*Boundaries: National Autonomy and its Limits*, ed. Peter G. Brown & Henry Shue, Rowman & Littlefield, 1981)

Lomborg, Bjorn (2001): *The Skeptical Environmentalist: Measuring the Real State of the World* (Cambridge University Press)

Macdonald, Graham and Pettit, Philip (1981): *Semantics & Social Science* (Routledge & Kegan Paul)

May, Larry (1987): *The Morality of Groups: Collective Responsibility, Group-Based Harm, and Corporate Rights* (University of Notre Dame Press)

May, Larry (1992): *Sharing Responsibility* (University of Chicago Press)

Mill, John Stuart (1859): *On Liberty* (*On Liberty and Other Essays*, J S Mill, ed. John Gray, OUP, 1991)

Mill, John Stuart (1861): *Utilitarianism* (*On Liberty and Other Essays*, J S Mill, ed. John Gray, OUP, 1991)

Miller, David (2004): "Holding Nations Responsible" (*Ethics* 114)

Moellendorf, Darrel (2002): *Cosmopolitan Justice* (Westview Press)

Murphy, Liam (1993): "The Demands of Beneficence" (*Philosophy and Public Affairs* 22)

Nozick, Robert (1974): *Anarchy, State and Utopia* (Basic Books)

Nussbaum, Martha (1990): "Aristotelian Social Democracy" (*Liberalism and the Good*, ed. R. Bruce Douglass, Gerald M. Mara & Henry S. Richardson, Routledge, 1990)

Nussbaum, Martha (1998): "The Good as Discipline, the Good as Freedom" (*Ethics of Consumption*, ed. Crocker and Linden, Rowman & Littlefield, 1998)

Nussbaum, Martha (2000a): "Aristotle, Politics, and Human Capabilities: A Response to Antony, Arneson, Charlesworth, and Mulgan" (*Ethics* 111)

Nussbaum, Martha (2000b): *Women and Human Development* (Cambridge University Press)

Oreskes, Naomi (2004): "The Scientific Consensus on Climate Change" (*Science* 306)



- Otsuka, Michael (1991): "The Paradox of Group Beneficence" (*Philosophy and Public Affairs* 20)
- Papineau, David (1978): *For Science in the Social Science* (Macmillan)
- Parfit, Derek (1984): *Reasons and Persons* (Oxford University Press, 1987 corrected reprint)
- Perry, Stephen (1992): "The Moral Foundations of Tort Law" (*Iowa Law Review* 77)
- Pettit, Philip (2001): "Deliberative Democracy and the Discursive Dilemma" (*Philosophical Issues* 11, ed. Ernest Sosa & Enrique Villanueva)
- Pierik, Roland, & Robeyns, Ingrid (2007): "Resources versus Capabilities: Social Endowments in Egalitarian Theory" (*Political Studies* 55)
- Pogge, Thomas (2002): *World Poverty and Human Rights* (Polity Press)
- Pogge, Thomas (2004): "Can the Capability Approach be Justified?" (*Global Inequalities*, ed. Nussbaum & Flanders, *Philosophical Topics* 30 (special issue))
- Rawls, John (1971): *A Theory of Justice* (Oxford University Press, revised edition 1999)
- Rawls, John (1993): *Political Liberalism* (Columbia University Press, paperback edition 1996)
- Rawls, John (1999): *The Law of Peoples* (Harvard University Press)
- Raz, Joseph (1986): *The Morality of Freedom* (Oxford University Press)
- Regan, Donald (1980): *Utilitarianism and Co-operation* (Oxford University Press)
- Robeyns, Ingrid (2005): "The Capability Approach: A Theoretical Survey" (*Journal of Human Development* 6)
- Robinson, Paul H (1975): "A Theory of Justification" (*Justification and Excuse in the Criminal Law*, ed. Michael Louis Corrado, Garland Publishing, 1994)
- Rousseau, Jean Jacques (1755): *A Discourse on the Origin of Inequality (The Social Contract and Discourses*, Rousseau, Everyman, 1913, revised edition 1973)
- Ruben, David-Hillel (1985): *The Metaphysics of the Social World* (Routledge & Kegan Paul)
- Runciman, W G, and Sen, Amartya K (1965): "Games, Justice and the General Will" (*Mind* 74)
- Runciman, W G, and Sen, Amartya K (1974): "Prisoner's Dilemma and Social Justice: A Reply" (*Mind* 83)

- Scheffler, Samuel (2001): *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought* (Oxford University Press)
- Sen, Amartya (1980): "Equality of What?" (*Liberty, Equality and Law: Selected Tanner Lectures on Moral Philosophy*, ed. Sterling M. McMurrin, Cambridge University Press 1987)
- Sen, Amartya (1992): *Inequality Reexamined* (Oxford University Press)
- Sen, Amartya (1999): *Development as Freedom* (Oxford University Press)
- Sheehy, Paul (2001): *The Ontological and Moral Status of Social Groups* (Kings College London PhD thesis)
- Sheehy, Paul (2002): "On Plural Subject Theory" (*Journal of Social Philosophy* 33)
- Shrader-Frechette, Kristin (2002): *Environmental Justice: Creating Equality, Reclaiming Democracy* (Oxford University Press)
- Shue, Henry (1980): *Basic Rights* (Princeton University Press, 2<sup>nd</sup> edition, 1996)
- Shue, Henry (1992): "The Unavoidability of Justice" (*The International Politics of the Environment*, ed. Andrew Hurrell and Benedict Kingsbury, Oxford University Press, 1992)
- Shue, Henry (1993): "Subsistence Emissions and Luxury Emissions" (*Law and Policy* 15)
- Shue, Henry (1999): "Bequeathing Hazards: Security Rights and Property Rights of Future Humans" (*Global Environmental Economics: Equity and the Limits to Markets*, ed. Mohammed Dore and Timothy Mount, Blackwell, 1999)
- Singer, Peter (1972): "Famine, Affluence, and Morality" (*Philosophy and Public Affairs* 1)
- Singer, Peter (2002): *One World: The Ethics of Globalization* (Yale University Press)
- Smyth, John (1972): "The Prisoner's Dilemma II" (*Mind* 81)
- Steiner, Hillel (1996): "Duty-Free Zones" (*Proceedings of the Aristotelian Society*, New Series 96)
- Tuomela, Raimo (2002): *The Philosophy of Social Practices* (Oxford University Press)
- US Global Change Research Program (2000): *Climate Change Impacts on the United States: The Potential Consequences of Climate Variability and Change* (Cambridge University Press; available from [www.usgcrp.gov](http://www.usgcrp.gov))

Waldron, Jeremy (1981): "A Right to Do Wrong" (*Ethics* 92)

Weinberg, Rivka (2002): *Procreative Justice: A Contractualist Account* (*Public Affairs Quarterly* 16)

Wiggins, David (2001): *Sameness and Substance Renewed* (Cambridge University Press)

Williams, Andrew (2002): "Dworkin on Capabilities" (*Ethics* 113)

Wolff, Jonathan and De-Shalit, Avner (2007): *Disadvantage* (Oxford University Press)